

La humanidad se encuentra en un momento crucial de su historia. La masa de datos adquiridos es formidable. Se necesitan nuevos instrumentos para simplificarlos, condensarlos o la inteligencia no podrá nunca remontar las dificultades que la aplastan, ni realizar los progresos que se vislumbran y a los que aspira (Otlet, 1934, p. 430)

RESUMEN (ESPAÑOL)

Buena parte de la información político-económica es difundida por los medios de comunicación social a través de noticias en formato digital. Sin embargo, dichos contenidos, y los de cualquier tipo en general, no tienen una estructura homogénea y se publican en grandes cantidades, lo que dificulta la extracción de conocimiento formalizado. Para resolver estos problemas a lo largo del tiempo se han desarrollado herramientas de organización y representación de la información. En los últimos años, en el marco de la web semántica aparecieron las ontologías, vocabularios que pueden mejorar la representación del contenido de documentos con un bajo nivel de estructuración, incorporando nuevos elementos a las herramientas tradicionalmente utilizadas. Ahora bien, el modelado de ontologías es complejo y requiere, entre otros, del uso de técnicas de procesamiento del lenguaje natural (indización automática) para dotar al vocabulario de elementos para la descripción de contenidos.

En este sentido, se plantea como objetivo general el modelado de una ontología con la que contribuir a la representación de acciones del dominio político-económico y facilitar la comprensión de los acontecimientos del mundo real en dicho contexto. Se propone la hipótesis de si el empleo de noticias de prensa digital como fuente para la adquisición de conocimiento y la aplicación de técnicas de indización humana y semiautomática para la extracción de términos son adecuados para el modelado de la ontología propuesta.

Para ello, se sigue una metodología que comprende la configuración de una muestra de noticias del dominio político-económico de diarios generalistas de ámbito internacional; el análisis de la teoría de la Gramática de Casos aplicado al modelado de una ontología general para la descripción de acciones (ONA); la indización humana y semiautomática de una submuestra de noticias para la modificación de ONA y el modelado inicial de la ontología de dominio ONAPE (ONTología de Acciones en Política-Economía); y el mapeado inicial con otros vocabularios, orientado a la identificación de elementos equivalentes y a la definición de clases y propiedades. ONAPE es evaluada a partir de la instanciación de las palabras clave extraídas del código fuente de noticias de una submuestra con un software desarrollado *ad hoc*, MetadadosHTML y la anotación semántica de un recorte de esta submuestra. Finalmente, se analiza la adecuación de ONAPE (y de ONA) al dominio objeto de estudio, mediante la aplicación de fórmulas de precisión, exhaustividad y medida-F. En todos los casos se obtienen valores superiores a 0.90, lo que asegura la corrección y especificación de los elementos de la ontología modelada.

La metodología utilizada se ha revelado útil para el propósito definido la metodología empleada. Por tanto, la hipótesis se cumple, es posible modelar una ontología de dominio con base en un componente teórico asentado, la Gramática de Casos y se atiende el objetivo planteado. Sin embargo, para que ONA y ONAPE sigan siendo útiles es necesaria su actualización por medio del análisis de nuevos *corpora* de noticias. En lo referente a MetadadosHTML, se pone en evidencia la dificultad del intercambio de información entre sistemas, ya que, aunque existen numerosos esquemas de metadatos para la descripción de noticias, ninguno es un estándar claro. Como trabajos futuros se plantean, entre otros: el uso de ONA u ONAPE en proyectos orientados al aprendizaje automático de software para la descripción automática de documentos y la especialización de ONAPE en subdominios concretos.

PALABRAS CLAVE

dominio política-economía, Gramática de Casos, ontologías de dominio, ONA, ONAPE, MetadadosHTML, metadatos, modelado de ontologías, web semántica, modelado de acciones

RESUMO (PORTUGUÊS)

Uma grande parte da informação político-econômica é difundida e conhecida através dos jornais digitais, por meio de notícias. Não obstante, estes conteúdos, e os conteúdos de qualquer tipo de documento em geral, não têm uma estrutura homogênea. Por outro lado, o volume de informação disponibilizada em Internet é enorme. Estas duas questões dificultam a extração de conhecimento formalizado. Para resolver os problemas indicados, tem sido desenvolvidas ferramentas de organização e representação da informação. No marco da web semântica, as ontologias aportam uma solução específica. Estas podem melhorar a representação do conteúdo de documentos com um nível baixo de estruturação diante da incorporação de elementos novos às ferramentas tradicionalmente utilizadas. Contudo, a modelagem de ontologias é complexa e requer, entre outros, do uso de técnicas de processamento da linguagem natural (indexação automática) para dotar ao vocabulário de elementos com vistas à descrição de conteúdos.

Neste sentido esta pesquisa visa à modelagem de uma ontologia que possa contribuir para a representação de ações do domínio político-econômico e facilitar a compreensão dos acontecimentos do mundo real neste contexto. Propõe-se a hipótese de se o uso de notícias de jornais digitais como fonte para a aquisição de conhecimento, por um lado, e a aplicação de técnicas de indexação humana e semiautomáticas para a obtenção de términos, por outro, são adequados para a posterior modelagem da ontologia proposta.

Para isto, segue-se uma metodologia que compreende a configuração de uma amostra de notícias do domínio político-econômico de jornais generalistas de âmbito internacional; a análise da teoria da Gramática de Casos aplicada à modelagem de uma ontologia geral para a descrição de ações, ONA; a indexação humana e semiautomática de uma sub-amostra de notícias para a modificação de ONA e a modelagem inicial da ontologia de domínio ONAPE (ONtología de Acción de Política-Economía); e o mapeamento inicial com outros vocabulários, orientado à identificação de elementos equivalentes e à definição de classes e propriedades. ONAPE é avaliada a partir da instanciação das palavras-chave das notícias de uma sub-amostra extraídas do seu código fonte com o software *ad hoc*, denominado MetadadosHTML, e a anotação semântica dum recorte desta sub-amostra. Finalmente, analisa-se a adequação de ONAPE ao domínio objeto de estudo por meio da aplicação de fórmulas de precisão e exaustividade. Para todos os casos se obtêm valores superiores a 0.90, o que assegura a correção e especificação dos elementos da ontologia modelada.

A metodologia utilizada também tem-se revelada útil para o propósito desta pesquisa. Portanto, a hipótese se cumpre, é possível modelar uma ontologia do domínio política-economía com base em um componente teórico assentado, a Gramática de Casos e com técnicas de indexação humana e automática. No entanto, para que ONA e ONAPE continuem a ser úteis e preciso aprofundar na sua avaliação, mesmo como na sua atualização por meio da análise de novos textos. No que se refere ao MetadadosHTML, foi útil para o propósito definido, mas tornou-se evidente a dificuldade da troca de informação entre sistemas pois, embora existem numerosos esquemas de metadados para a descrição de notícias de atualidade, nenhum deles constitui um padrão claro. Como trabalhos futuros tem sido levantados, entre outros, o uso de ONA ou ONAPE em projetos orientados à aprendizagem automática de software para a descrição automática de documentos e a especialização de ONAPE em subdomínios concretos.

palavras-chave

domínio política-economía, Gramática de Casos, ontologias de domínio, ONA, ONAPE, MetadadosHTML, metadados, modelagem de ontologias, web semântica, modelagem de ações

ABSTRACT (ENGLISH)

A huge amount of political-economical information is disseminated through digital social media by pieces of news. However, these contents and, in general terms, any type of information do not have a homogeneous structure and are published in large amounts. This makes the extraction of formal knowledge difficult. In order to resolve the problems generated by these issues, knowledge organisation and representation tools have been created. The semantic web and specifically the ontologies can improve the representation of the content in documents with a low level of structuring, by adding new elements to the tools traditionally used. But the ontologies-modelling is complex and requires the use of natural language processing techniques -automatic indexation-. In this way, a designed methodology can provide the vocabulary with elements that can describe contents.

In this sense, the general aim of this thesis is modelling an ontology that can contribute to the representation of actions in the political-economical domain and can facilitate the understanding of the real world facts in this context. The hypothesis is if digital press pieces of news, used as a source of knowledge acquisition, and the application of both human indexation and semi-automatic indexation techniques -for extraction of terms- are adequate for modelling the proposed ontology.

For this purpose, a methodology has been defined and comprises the following steps: a sample and subsample construction composed of pieces of news of the political-economical domain. This is based in a selection of digital international newspapers; the Grammar Case analysis and its application to the modelling of a general ontology for the description of actions (ONA); both human indexing and semi-automatic indexing to the subsample₁ in order to modify ONA and to model a domain ontology for the description of political-economical actions (ONAPE); and the initial mapping with some vocabularies, in order to identify equivalent elements and define classes and properties. ONAPE (and ONA) are evaluated by instantiating ONAPE with some keywords (extracted from source codes of the subsample₂ with MetadadosHTML) and by the semantic annotation of a cut of this subsample. Finally, the adaptation of ONAPE (and thus of ONA) to the studied domain is analysed by the application of both accuracy and recall equations. In all cases higher than 0.9 values are obtained. This ensures the correctness and specification of the elements of the modelled ontology.

The designed methodology has proved profitable for the purpose of this thesis. With the Grammar Case as its main theoretical component, the methodology can be used for the modelling of others domain ontologies. However, for these tools remaining useful they need to be updated by analysing new *corpora*. In relation to MetadadosHTML, there is evidence of the difficulty of the information interchange. This has to do with the fact that there are numerous metadata schemas for the description of pieces of news, but none of them has become a standard. Finally, future works are proposed, such as the use of both ONA and ONAPE in projects focused on the automatic learning software for the automatic description of documents and the specialization of ONAPE in specific subdomains.

KEYWORDS

political-economical domain, case grammar, domain ontologies, ONA, ONAPE, MetadadosHTML, metadata, ontologies modelling, semantic web, actions modelling

AGRADECIMIENTOS

Me gustaría comenzar esta sección dando las gracias a mis directores, los doctores **Juan Antonio Pastor Sánchez** y **Rodrigo Martínez Béjar**, por su participación en esta investigación y darme la oportunidad de conocer el mundo académico. Mi paso por la Universidad de Murcia, las estancias en Portugal y Brasil, los encuentros con amigas y amigos, profesores e investigadores de diferentes universidades han dado como fruto este trabajo. A todas estas personas también agradezco su ayuda y disposición. Detallo aquí, por orden alfabético, la participación de algunas:

Ana Alice Baptista, Universidade do Minho (UM). Minha orientadora no sanduíche na Escola de Engenharia, Departamento de Sistemas de Informação. Foi com você que conheci os esquemas de metadados e recebi as primeiras dicas sobre como pesquisar.

Dagobert Soergel, University of Buffalo, thanks for your advices. I learnt some useful aspects about ontologies, how to model them and English.

Daniela da Silva, Universidade Federal do Espírito Santo (UFES), você me recebeu na sua casa sem me conhecer, lá em Vitória. Naqueles dias você me aconselhou, prestou-me seus livros e artigos e até a sua tese. Essa pesquisa tão legal foi muito boa para mim.

(Seu) **Edberto Fereda**, Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), meu orientador no sanduíche em Marília. Você sofreu lendo parte de este documento, revisou ele e me deu ideias para melhorar minha pesquisa. Ele também me prestou um monte de livros básicos de Ciência da Informação e sorriu até quando a gente não concordava.

Eduardo Ribeiro Felipe, Universidade Federal de Minas Gerais (UFMG), amigo, futuro doutor. Ele criou para mim o aplicativo desenvolvido nesta pesquisa, MetadadosHTML para extrair metadados do código fonte e revisou o resumo em português para mim. Ele resistiu minhas queixas sobre a situação da pesquisa na Espanha, a educação e política nacional, falamos da vida, de gatos e do tempo que nem chegou.

Francisco Javier Martínez Méndez, Universidad de Murcia, siempre con buena disposición para echar una mano cuando ha hecho falta y que con aquel “te vas a ir a Portugal” me envió a dos países, una nueva lengua (o dos) y un montón de gente bonita.

Gercina Lima, Universidade Federal de Minas Gerais (UFMG), orientadora em Belo Horizonte, quase cuidou de mim como da sua filha. Ela também me recebeu na sua casa e me ajudou com conselhos e livros.

Giancarlo Guizzardi, Universidade Federal do Espírito Santo (UFES). Conheci você na Escola de Verão de Ontologías, em Vitória. Você me encorajou para continuar pesquisando sobre ontologías e me falou da importância dos documentalistas no modelado e construção delas.

Ivo Pierozzi Junior, Embrapa Informática Agropecuária (Campinas, São Paulo, Brasil), que mostrou para mim o aplicativo e-Termos, que depois utilizei nesta pesquisa. Ele também escutou o que eu achava naquela época que ia fazer e me explicou como poderia utilizar e-Termos na tese.

José Ramón Sánchez Gallego (Yosi), University of Kentucky, no quiero ni pensar en el follón que te he dado, desde el principio y hasta el final, con dudas sobre el doctorado, cómo escribir, qué hacer después, etc. Has sido una de las personas gracias a las cuáles no he abandonado. Todavía no tengo claro si tengo que darle las gracias o matarlo.

Juana María Ruiz Martínez, doctora en Documentación por la Universidad de Murcia, compañera de laboratorio, consejera y amiga. Otra amiga que ha pasado por el calvario de leer parte de la tesis, me ha prestado bibliografía, me ha aconsejado y que me ha ayudado.

Mari Carmen Legaz García, Universidad de Murcia, sufriste de primera mano la aplicación de la Gramática de casos de Fillmore en el modelado inicial de ONA. La Auténtica dedicó varias horas de su vida, sin matarme, a explicarme conceptos y relaciones en ontologías. Sin ella, el resultado habría sido mucho peor.

Manuel Quesada Martínez, Universidad de Murcia. Tuviste la santa paciencia de leer uno de los artículos sobre análisis de metadatos en códigos fuente :P

Maurício Almeida, Universidade Federal de Minas Gerais, me recebeu na sua oficina e escutou meu projeto de tese. Me aconselhou e enviou para mim alguns documentos que depois utilicei nesta pesquisa. Foi ele que mostrou para mim um rascunho básico do que a tese poderia ser.

Michael Uschold, Semantic Arts, Inc., I met you at the Second Interdisciplinary Summer School on Ontological Analysis. You cheered me up when I was very depressed due to my poor knowledge about ontologies. In your speech, and I think because of me, you spoke about how and why the documentalists are important during an ontology modeling.

Morgana Carneiro de Andrade, Universidade do Minho, mais uma amiga com quem compartilhei apartamento e quarto em Guimarães, mas também conhecimentos e desespero :P Também ela me ajudou com o resumo em português. A gente conversou muito, se visitou e ajudou. Vitória está longe, mas muito perto.

Pilar Alonso Lifante, compañera de doctorado en Documentación, y también Javi. Con vosotros también compartí parte de los desvelos, inseguridades pero, sobre todo, ideas. Esta pareja es una mina pendiente de ser explotada.

Raquel Lorente Pallarés, Universidad de Murcia, si no hubiera sido por ti, ni habría optado a la beca de Doctorado. Siempre has estado ahí, desde el inicio, desde el día que llegué con mi padre a la Facultad para matricularme por primera vez en la universidad. Para cada problema, Raquel tiene una solución, si no existe, Raquel la crea :D

Rocío Pintado Navarro, Universidad de Murcia, espadachina de las letras y la corrección, amante de la RAE y de las comas, revisora de textos en español y portugués, consultora de términos judiciales y, sobre todo, amiga. Y ojo, de las mejorcicas. Desde que nos conocimos en el máster, anda que no hemos cosas tú y yo.

Zaira Regina Zafalon, Universidade Federal de São Carlos, São Paulo (Brasil) me recebeu na sua casa, organizou para mim uma estadia muito legal, e corrigiu para mim objetivos e estrutura da tese num momento que já estava avançada. Ela me explicou os tipos de pesquisa e me encorajou.

Aún hay personas que, de una u otra forma, me han acompañado durante este proceso. Con algunas empecé Biblioteconomía y Documentación, allá por 2001, en la entonces Facultad de Documentación. A otras las conocí en Publicidad y Relaciones Públicas, en 2004. Después llegó la gente de la Licenciatura de Documentación, que acabé en 2011. Al final, el Máster de Gestión de Información en las Organizaciones, en 2013. Con otras he compartido espacio, penas y alegrías, quejas y desvaríos en la Facultad de Informática, en el chiringuito de la Facultad de Comunicación y Documentación, en la Biblioteca Regional de Murcia o de paseo por algún pueblo de esta nuestra Región, que no todo va a ser estudiar.

Sería imposible establecer un orden de preferencia, son muchas y muy queridas. Así que el orden es aleatorio. Vamos con ello: Piedad Fernández Toledo, Astrid Duque Ramos (Alejandro, David y el que está por venir), la sección internacional (Omar Colombator, Mario Andrés Paredes, Pilar Salas Zarate), el ciudadano Ángel, Teddy Miranda Mena, Lucía Serrano Luján (Luly), Miguel Ángel Rodríguez, Ginés Almagro, el desaparecido Fran Collado, Fran Abad y su Pedro, Ceren Abacioglu (my little sister), Ciliana Regina Colombo, Almudena Martínez Díaz (Jarmu), Sophia Piacenza, Manolo Hernández Pérez (en su atalaya de Mordor), Dani Megías (sufridor de la web 2.0.), la hermana Lupe Taberner, Miguel Ángel Cárceles Parra, Mariángeles (con B de bonita), la funcionaria y artista María García Hernández, Andrea, Lucía Martínez López (Lu), Rodrigo Cuello, Rosa Fernández Cabadas (qué follón te he dado), Pablo Baños López (mi primo pelirrojo), Pedro Jesús Sánchez, Eloy, Mariano Almagro, Jorge Ferrando, Ana Arce, José Luis Rodríguez. Además, en los últimos meses han aparecido otras personas, la gente de Odilo que, pobre, ha sufrido la tesis a su manera: Isa, Jorge, Valentín (sobre todo), María, Andreh, Nacho, Javi, Leo, Juan Luis y muchas más.

Quisiera también destacar la importancia de las bibliotecas que me han acogido durante estos años y no sólo como salas de estudio. La Biblioteca Regional de Murcia, la Biblioteca de Puertas de Castilla, la Biblioteca Municipal de Alguazas, la Biblioteca del Cuartel de Artillería, la Biblioteca General de la Universidad de Murcia y la Biblioteca de Molina de Segura, entre otras. Han sido, muchas veces, los sitios en los que he trabajado, he descansado, he consultado fuentes de información (en papel y en línea), he tomado libros en préstamo o los he leído allí mismo. Nunca las valoramos lo suficiente.

Por último quería agradecer el apoyo de mi familia, cada uno/a a su manera. A veces, con un simple café en un bar, con Javi y Raquel, o comer en el campo es suficiente para desconectar. Y, por fin, el mayor agradecimiento es para Simón. Que ha sido la segunda persona que más ha sufrido con esta tesis. Puede que hasta merezca aparecer en la portada. Ha aguantado estoicamente cada queja, cada problema y cada enfado pero siempre, SIEMPRE, ha estado ahí. No creo que pudiera encontrar una persona mejor para mi. Soninoni.

Este trabajo ha sido posible gracias a la financiación de la Universidad de Murcia a través del Proyecto 15970 de Contratos Predoctorales UM y resoluciones de convocatoria, concesión y renovación de la Beca-Contrato predoctoral y de estancias breves: Resolución R.-406/2011 de 23 de Septiembre, del Rector de la Universidad de Murcia por la que se aprueban las Bases Regulatoras y la convocatoria de Becas-Contratos Predoctorales (FPU) de la Universidad de Murcia; Resolución del Vicerrectorado de Investigación y de Campus de Excelencia de la Universidad de Murcia de 26 de diciembre de 2011 de la convocatoria de Becas-Contrato Predoctorales para 2012; Resoluciones R.-779/2012 de 11 de diciembre, R.-1050/2013 de 20 de diciembre y R.-930/2014 de 17 de diciembre del Rector de la Universidad de Murcia por la que se renuevan las ayudas concedidas a los beneficiarios de la Convocatoria de Becas-Contrato Predoctorales (FPU) de la Universidad de Murcia para los años 2013, 2014, 2015-16; Resolución de 15 de abril de 2013 del Vicerrector de Investigación e Internacionalización de la Universidad de Murcia de concesión de ayudas para estancias en España o en el extranjero para beneficiarios del Programa de Becas-Contrato Predoctorales de la Universidad de Murcia; Resoluciones del Rector R.-365/2014 de 8 de mayo y R.-369/2016 de 13 de mayo de la Universidad de Murcia por la que se resuelve con carácter definitivo la convocatoria de Ayudas para estancias en el extranjero para beneficiarios del Programa de Contratos Predoctorales (FPU) de la Universidad de Murcia para 2016.

*

**

SUMARIO

1. INTRODUCCIÓN.....	1
1.1. OBJETIVOS.....	7
1.2. HIPÓTESIS DE TRABAJO.....	8
1.3. MARCO DE REFERENCIA UNESCO.....	8
1.4. ESTRUCTURA DEL TRABAJO.....	8
2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA.....	11
2.1. LA WEB SEMÁNTICA Y OWL.....	23
2.2. ONTOLOGÍAS. CONCEPTO Y CARACTERÍSTICAS.....	29
2.2.1. COMPONENTES DE LA ONTOLOGÍA.....	31
2.2.2. TIPOS DE ONTOLOGÍAS.....	37
2.2.3. METODOLOGÍAS PARA CONSTRUIR Y EVALUAR ONTOLOGÍAS. . .	40
2.2.4. HERRAMIENTAS PARA ONTOLOGÍAS: PROTÉGÉ.....	45
2.2.5. REPOSITORIOS DE VOCABULARIOS: <i>LINKED OPEN VOCABULARIES</i>	49
2.2.6. REUTILIZACIÓN DE VOCABULARIOS: MAPEO.....	52
2.3. LA INDIZACIÓN.....	55
2.3.1. INDIZACIÓN HUMANA E INDIZACIÓN AUTOMÁTICA.....	56
2.3.2. EXTRACCIÓN DE TÉRMINOS CON N-GRAMAS: e-TERMOS. ANÁLISIS MORFOLÓGICO: GRAMPAL.....	61
2.3.3. INDIZACIÓN DE NOTICIAS. LA GRAMÁTICA DE CASOS.....	64
2.3.4. EVALUACIÓN DE LA INDIZACIÓN.....	67
3. PROCEDIMIENTOS METODOLÓGICOS.....	69
3.1. ASPECTOS METODOLÓGICOS PREVIOS.....	70
3.2. TIPO DE INVESTIGACIÓN.....	73
3.3. LIMITACIONES DE LA INVESTIGACIÓN.....	75
3.4. DESARROLLO DE LA INVESTIGACIÓN.....	76
3.5. APLICACIONES Y PÚBLICO OBJETIVO.....	81
4. RESULTADOS.....	87
4.1. CONFIGURACIÓN DE LA MUESTRA.....	89
4.1.1. LA NOTICIA, FUENTE DE TÉRMINOS PARA LA REPRESENTACIÓN DEL CONOCIMIENTO.....	89
4.1.2. SELECCIÓN DE PERIÓDICOS Y NOTICIAS.....	91
4.1.3. DEFINICIÓN DE SUBMUESTRAS.....	94
4.2. MODELADO DE LA ONTOLOGÍA DE ACCIONES ONA BASADO EN LA GRAMÁTICA DE CASOS.....	95
4.2.1. ANÁLISIS DEL MARCO TEÓRICO Y ENTIDADES BÁSICAS.....	95
4.2.2. MODELADO INICIAL DE ONA.....	101
4.2.3. DATOS ESTADÍSTICOS DEL PROCESO.....	110

4.3. DESCUBRIMIENTO DE CONOCIMIENTO POR INDIZACIÓN. MODELADO DE LAS ONTOLOGÍAS.....	111
4.3.1. ATRIBUCIÓN DE TÉRMINOS POR INDIZACIÓN HUMANA.....	111
4.3.2. EXTRACCIÓN DE TÉRMINOS POR INDIZACIÓN AUTOMÁTICA. .	112
4.3.3. MODIFICACIÓN DE ONA.....	113
4.3.4. MODELADO INICIAL DE ONAPE A PARTIR DE LOS RESULTADOS DE LA INDIZACIÓN HUMANA.....	114
4.3.5. MODIFICACIÓN DE ONAPE A PARTIR DE LA INDIZACIÓN AUTOMÁTICA.....	118
4.3.6. DATOS ESTADÍSTICOS DEL PROCESO.....	120
4.4. REUTILIZACIÓN DE ELEMENTOS PREVIOS: MAPEO CON OTROS VOCABULARIOS.....	123
4.4.1. PRESELECCIÓN DE VOCABULARIOS.....	124
4.4.2. SELECCIÓN DE VOCABULARIOS.....	139
4.4.3. MAPEO DE VOCABULARIOS.....	141
4.5. EVALUACIÓN DE LAS ONTOLOGÍAS ONA Y ONAPE.....	149
4.5.1. PREPROCESAMIENTO DE LA SUBMUESTRA ₂ PARA LA EVALUACIÓN DE ONAPE: METADADOSHTML.....	150
4.5.2. IDENTIFICACIÓN DE RELACIONES SEMÁNTICAS NOTICIA-ONAPE PARA LA VALIDACIÓN DE ONAPE.....	157
4.5.3. VALIDACIÓN DE ONAPE MEDIANTE INSTANCIACIÓN.....	160
4.5.4. VALIDACIÓN POR ANOTACIÓN SEMÁNTICA.....	174
4.6. ESPECIFICACIONES DE LAS ONTOLOGÍAS.....	180
5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES.....	181
5.1. CONCLUSIONES.....	183
5.2. TRABAJOS FUTUROS.....	187
5.3. PUBLICACIONES CIENTÍFICAS.....	191
RESUMO EM PORTUGUÊS.....	197
1. INTRODUÇÃO.....	199
1.1. OBJETIVOS.....	201
1.2. HIPÓTESE DE TRABALHO.....	201
2. FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA.....	202
2.1. AS ONTOLOGIAS.....	202
2.2. A INDEXAÇÃO.....	206
3. PROCEDIMENTOS METODOLÓGICOS E RESULTADOS.....	210
3.1. CONSTRUÇÃO DA AMOSTRA.....	210
3.2. DESCOBRIMENTO DE CONHECIMENTO.....	216
3.3. MODELAGEM DE ONAPE A PARTIR DOS RESULTADOS DA INDEXAÇÃO HUMANA.....	218
3.4. MAPEAMENTO COM OUTROS VOCABULÁRIOS.....	220
3.5. AVALIAÇÃO DE ONA E ONAPE.....	222
3.6. ESPECIFICAÇÃO DE ONA E ONAPE.....	227

4. CONCLUSÕES E PROPOSTAS.....	227
4.1. CONCLUSÕES.....	227
4.2. PROPOSTAS PARA O FUTURO.....	232
BIBLIOGRAFÍA.....	237
ANEXOS.....	271
A. DESCRIPCIÓN DE VOCABULARIOS MAPEADOS.....	273
A.1. HERRAMIENTAS REUTILIZADAS POR LOS VOCABULARIOS.....	320
A.2. ENLACES DE INTERÉS DE LOS VOCABULARIOS.....	321
B. ESPECIFICACIÓN DE ONA.....	322
B1. INTRODUCCIÓN A LA ESPECIFICACIÓN.....	322
B2. RESUMEN.....	323
B3. ELEMENTOS DE ONA.....	324
C. MARCADO SEMÁNTICO DE NOTICIAS DEL TEMA 1.....	340

ÍNDICE DE FIGURAS

FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Figura 1: Continuo de los Lenguajes Documentales (Codina y Pedraza-Jiménez, 2015).....	14
Figura 2: Ficha de Cervantes en catálogo de Autoridades de la Biblioteca Nacional de España.....	15
Figura 3: Descripción del término "sistemas políticos" en Newscodes.....	17
Figura 4: Ficha en EUROVOC del descriptor "régimen político".....	19
Figura 5: Algunos elementos de la ontología Event en Protégé.....	21
Figura 6: Descripción de una noticia con metadatos.....	22
Figura 7: Metadato título en el esquema Dublin Core.....	23
Figura 8: Arquitectura de la web semántica.....	24
Figura 9: Gráfico informal de tripletas ejemplo.....	27
Figura 10: Triángulo semiótico de Ogden y Richards (1923).....	31
Figura 11: Dominio y rango de la relación o propiedad tipo objeto "ejecuta".....	32
Figura 12: Captura de pantalla de Protégé 5.0.0.....	46
Figura 13: Pestaña para la definición de clases en Protégé 5.0.0.....	47
Figura 14: Pestaña para la definición de propiedades en Protégé 5.0.0.....	48
Figura 15: Vocabularios en Linked Open Vocabularies según su grado de reutilización.....	50
Figura 16: Datos recogidos por LOV para el vocabulario FOAF.....	51
Figura 17: Patrón de equivalencia exacta entre clases de dos ontologías.....	54
Figura 18: Alineamiento de la propiedad name de vCard y FOAF.....	54
Figura 19: Relación entre descripción de documentos-indización-modelado de vocabularios.....	60
Figura 20: Comparación de términos con n-gramas.....	61
Figura 21: Ejemplo de uso de Grampal.....	63
Figura 22: Paradigma de Lasswell y estructura de la noticia.....	64
Figura 23: Relación Lasswell, Fillmore y García Gutiérrez. Basada en García Gutiérrez (2014)....	66
Figura 24: Fórmula para el cálculo de la exhaustividad en una búsqueda.....	68
Figura 25: Fórmula para el cálculo de la precisión de una búsqueda.....	68
Figura 26: Fórmula para el cálculo de la medida-F en una búsqueda.....	68

PROCEDIMIENTOS METODOLÓGICOS

Figura 27: Tipologías documentales contempladas por ONAPE.....	81
Figura 28: Hipotético sistema de recuperación de información basado en ONAPE.....	83

RESULTADOS

Figura 29. División del mundo por zonas geopolíticas.....	92
Figura 30: Modelo ONA a partir de la interpretación literal de la Gramática de Casos.....	100
Figura 31: Relaciones básicas del modelo.....	101
Figura 32: a) Relaciones Agente-Accion e Instrumento-Accion.....	101
Figura 33: b) Relaciones Agente-Accion, Agente-Instrumento e Instrumento-Accion.....	102
Figura 34: c) Relaciones de una clase auxiliar con Agente, Instrumento y Accion.....	102
Figura 35: Relación entre AgenteAccionEjecutada y resto de clases relacionadas.....	103
Figura 36: Modelo ONA inicial.....	108
Figura 37: Modelo ONA inicial - Relaciones inversas.....	109
Figura 38: Relación esEjecutanteDe y las anotaciones "label" y "comment" en Protégé.....	110
Figura 39: Etapa 3 de e-Terms, listados de términos y visualización de lista de 1 grama.....	112
Figura 40: Relaciones partitivas y de antecendencia entre clases de ONA.....	114
Figura 41: Aplicación del enfoque middle-out en el modelo.....	115
Figura 42: Relaciones de sinonimia entre términos del modelo de ontología.....	116
Figura 43: Ejemplo de clase definida a partir de instancias extraídas por indización automática. .	119

Figura 44: Clases y relaciones de ONAPE a partir de Embajada de Estados Unidos en Yemen.....	119
Figura 45: Datos estadísticos a partir de la indización humana.....	121
Figura 46: Datos estadísticos a partir de indización humana + automática.....	122
Figura 47: Biography Vocabulario Timeline.....	129
Figura 48: Vista de la ontología TEO.....	130
Figura 49: Clases Instant e Interval de The Timeline Ontology.....	131
Figura 50: Elementos top level de SENSO.....	135
Figura 51: Mapeo de elementos del modelo de ontología y vocabularios analizados.....	141
Figura 52: Correspondencias tipos 1 y 3 de Sujeto y Agente en el modelo y DUL.....	142
Figura 53: Correspondencia de tipo 2 entre la clase Cosa del modelo y clases de FOAF.....	143
Figura 54: Equivalencia parcial entre la clase Agente del modelo y FOAF.....	144
Figura 55: Correspondencia tipo 4 entre Cosa del modelo y Abstract y Object de DUL.....	144
Figura 56: Correspondencia tipo 1 con una relación del modelo y una relación de PROLES.....	146
Figura 57: Correspondencia de tipo 2 entre relaciones del modelo y de la ontología DUL.....	146
Figura 58: Representación gráfica de las dos etapas de validación de ONAPE.....	149
Figura 59: Infraestructura organizacional de MetadadosHTML (v1).....	151
Figura 60: Interfaz de MetadadosHTML (v1).....	152
Figura 61: Interfaz de MetadadosHTML (v2).....	153
Figura 62: Fórmula para el cálculo de la exhaustividad de ONAPE.....	159
Figura 63: Fórmula para el cálculo de la precisión de ONAPE.....	159
Figura 64: Fórmula para el cálculo de la medida-F de ONAPE.....	159
Figura 65: Tipos de relación entre palabras clave y ONAPE y grado de importancia.....	159
Figura 66: Relación entre las instancias AJS y The Asahi Shimbun.....	161
Figura 67: Instanciación de ONAPE con la instancia Apartheid.....	164
Figura 68: Acción migrante ejecutada por inmigrantes ilegales.....	167
Figura 69: Relación entre Monedas y Estados para el tema Caso de Corrupción en Turquía.....	168
Figura 70: Relación entre Aire y Calidad y otros elementos.....	170
Figura 71: Relación de Plan Básico de Energía con elementos de ONAPE.....	172
Figura 72: Medias de tipos de relaciones entre palabras clave y elementos de ONAPE.....	173
Figura 73: Clases e instancias identificadas en las noticias del tema 1 de la submuestra2.....	178
Figura 74: Medias de tipos de relaciones entre elementos de noticias y ONAPE.....	179

CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Figura 75: Relación entre Accion y ModuladorDeAccion.....	188
---	-----

RESUMO EM PORTUGUÊS

Figura 76: Domínio e range da propriedade "executa".....	203
--	-----

ANEXOS

Figura 77: Modelo de Event.....	285
Figura 78: Latitud y longitud del planeta Tierra.....	292
Figura 79: Clases, relaciones y propiedades para "Russian Federation" en GEOP.....	294
Figura 80: Diagrama de la ontología de la BBC.....	297
Figura 81: Interrelaciones entre las clases de las (micro)ontologías SNAP.....	302
Figura 82: POWER aplicado a la instancia "José Sócrates" (Moreira et al., 2013).....	312
Figura 83: Modelo de la ontología Proles. Fuente: (Daquino et al., 2014).....	318

ÍNDICE DE TABLAS

FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Tabla 1: Relaciones asociativas entre descriptores de un tesoro.....	18
Tabla 2: Descripción de las capas de la arquitectura de la web semántica.....	26
Tabla 3: Tipos de relaciones entre clases y algunas de sus propiedades.....	33
Tabla 4: Ejemplos de metodologías de construcción de ontologías.....	41
Tabla 5: Principios de construcción de ontologías (Mendonça y Almeida, 2014).....	43
Tabla 6: Medidas para evaluar ontologías desde el punto de vista de su corrección.....	44
Tabla 7: Tipos de relaciones entre elementos a mapear y patrones de alineamiento.....	53
Tabla 8: Cualidades de la indización humana y automática.....	57
Tabla 9: Etapas de procesamiento de la información en e-Termos.....	62
Tabla 10: 12W de una noticia, según García Gutiérrez (2014).....	65

PROCEDIMIENTOS METODOLÓGICOS

Tabla 1: Parte de cursos y congresos a los que se ha asistido.....	71
Tabla 2: Estancias durante el periodo de investigación.....	72
Tabla 3: Cursos y pruebas de nivel de idiomas realizados durante el periodo de investigación.....	72
Tabla 4: Puestos de trabajo anteriores relacionados con la investigación realizada.....	73

RESULTADOS

Tabla 1: Elementos de categorización de noticias.....	90
Tabla 2: Periódicos seleccionados.....	93
Tabla 3: Temáticas de las noticias de la submuestra2.....	95
Tabla 4: Clasificación de casos en entidades básicas.....	96
Tabla 5: Equivalencias entre Casos y elementos de ONA.....	103
Tabla 6: Clases de ONA.....	104
Tabla 7: Relaciones iniciales de ONA.....	106
Tabla 8: Palabras clave de descripción de noticias obtenidas por indización humana.....	111
Tabla 9: Datos estadísticos iniciales obtenidos en e-Termos.....	112
Tabla 10: Resumen del análisis de términos extraídos con e-Termos.....	113
Tabla 11: Propiedades tipo objeto incorporadas a ONA.....	113
Tabla 12: Frecuencias n-gramas por clase de ONA para los términos revisados.....	118
Tabla 13: Acrónimos, URL y último acceso a las ontologías preseleccionadas.....	123
Tabla 14: Elementos de la ontología LODE.....	128
Tabla 15: Criterios de selección de ontologías para su reutilización.....	140
Tabla 16: Clases de los vocabularios analizados para su mapeo en el modelo.....	145
Tabla 17: Relaciones de los vocabularios analizados para su mapeo en el modelo.....	148
Tabla 18: Tablas que componen la base de datos de MetadatosHTML.....	151
Tabla 19: Noticias sobre la disputa territorial por las Islas Senkaku.....	154
Tabla 20: Noticias en las que aparece Mandela como palabra clave.....	155
Tabla 21: Noticias relativas al rechazo de inmigrantes en Reino Unido.....	156
Tabla 22: Noticias sobre caso de corrupción en Turquía.....	156
Tabla 23: Noticias sobre mercado energético.....	157
Tabla 24: Instanciación de la submuestra2 en ONAPE. Tema 1.....	161
Tabla 25: Instanciación de la submuestra2 en ONAPE. Tema 2.....	163
Tabla 26: Instanciación de la submuestra2 en ONAPE. Tema 3.....	166
Tabla 27: Instanciación de la submuestra2 en ONAPE. Tema 4.....	168
Tabla 28: Instanciación de la submuestra2 en ONAPE. Tema 5.....	170
Tabla 29: Total de relaciones entre la submuestra2 y elementos de ONAPE por tema.....	172

Tabla 30: Exhaustividad, Precisión y medida-F por temas analizados.....	173
Tabla 31: Exhaustividad, Precisión y medida-F por temas analizados.....	174
Tabla 32: Colores de marcado semántico para la identificación de entidades básicas en noticias..	174
Tabla 33: Total de relaciones entre elementos del tema 1 de la submuestra2 y ONAPE.....	177
Tabla 34: Relaciones entre términos de noticias y elementos de ONAPE por noticia.....	179
Tabla 35: Exhaustividad, Precisión y medida-F por temas analizados.....	180

CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Tabla 1: Información sobre el metadato "palabras clave" para MetadadosHTML.....	191
Tabla 2: Artículos publicados.....	193
Tabla 3: Contribuciones a congresos.....	195
Tabla 4: Capítulos de libro publicados durante el periodo de investigación.....	196
Tabla 5: Presentaciones de esta investigación.....	196

RESUMO EM PORTUGUÊS

Tabela 1: 12W de uma notícia, segundo García Gutiérrez.....	208
Tabela 2: Jornais selecionados.....	210
Tabela 3: Temáticas das notícias da subamostra2.....	211
Tabela 4: Equivalências entre Casos e elementos de ONA.....	214
Tabela 5: Classes de ONA.....	214
Tabela 6: Relações iniciais de ONA.....	216
Tabela 7: Dados estatísticos iniciais obtidos em e-Termos.....	217
Tabela 8: Propriedades tipo objeto incorporadas a ONA.....	217
Tabela 9: Frequências n-gramas por classe de ONA para os termos revisados.....	218
Tabela 10: Dados estatísticos das modelagens de ONA e ONAPE.....	220
Tabela 11: Total de relações entre a sub-amostra2 e elementos de ONAPE por tema.....	224
Tabela 12: Exaustividade, Precisão e medida-F por temas analisados.....	225
Tabela 13: Total de elementos do tema 1 da sub-amostra2 identificados em ONAPE.....	225
Tabela 14: Relações entre termos de notícias e elementos de ONAPE por notícia.....	226
Tabela 15: Exaustividade, Precisão e medida-F por temas analisados.....	227
Tabela 16: Informação sobre o metadado "palavras-chave" para MetadadosHTML.....	235

ANEXOS

Tabla 1: Ejemplos de descripción de elementos en los vocabularios analizados.....	273
Tabla 2: Algunas clases de schema.org.....	274
Tabla 3: Parte de las relaciones de schema.org.....	275
Tabla 4: Parte de las propiedades tipo dato de schema.org.....	276
Tabla 5: Vocabularios que están siendo mapeados con schema.org.....	277
Tabla 6: Algunas clases de la ontología DUL.....	279
Tabla 7: Algunas relaciones de la ontología DUL.....	280
Tabla 8: Propiedades tipo dato en DUL.....	281
Tabla 9: Herramientas reutilizadas en DUL.....	281
Tabla 10: Clases de la ontología EVENT.....	282
Tabla 11: Propiedades tipo objeto de la ontología EVENT.....	284
Tabla 12: Propiedades tipo dato de la ontología EVENT.....	284
Tabla 13: Vocabularios reutilizados en EVENT.....	285
Tabla 14: Clases de FOAF.....	286
Tabla 15: Relaciones en FOAF.....	288
Tabla 16: Propiedades tipo dato de FOAF.....	289
Tabla 17: Vocabularios reutilizados en FOAF.....	289
Tabla 18: Clases de la ontología TIME.....	290
Tabla 19: Relaciones de la ontología TIME.....	291
Tabla 20: Propiedades tipo dato de la ontología TIME.....	291

Tabla 21: Vocabularios reutilizados en TIME.....	292
Tabla 22: Clases de GEO.....	293
Tabla 23: Relaciones de GEO.....	293
Tabla 24: Propiedades de anotación de GEO.....	293
Tabla 25: Clases de GEOP.....	295
Tabla 26: Propiedades tipo objeto de GEOP.....	295
Tabla 27: Algunas propiedades tipo dato de la ontología GEOP.....	296
Tabla 28: Clases de la ontología BBCCORE.....	299
Tabla 29: Relaciones de BBCCORE.....	300
Tabla 30: Propiedades tipo dato de BBCCORE.....	301
Tabla 31: Vocabularios reutilizados en BBCCORE.....	301
Tabla 32: Ontologías que conforman SNAP ontologies.....	302
Tabla 33: Vocabularios reutilizados en el conjunto de las ontologías SNAP.....	303
Tabla 34: Clases de la ontología (SNAP) PNA.....	303
Tabla 35: Relaciones de (SNAP) PNA.....	304
Tabla 36: Propiedades tipo dato de (SNAP) PNA.....	304
Tabla 37: Clases de la ontología (SNAP) PNC.....	304
Tabla 38: Relaciones de (SNAP) PNC.....	305
Tabla 39: Clases de la ontología (SNAP) PNE.....	306
Tabla 40: Relaciones de la ontología (SNAP) PNE.....	306
Tabla 41: Propiedades tipo dato de (SNAP) PNE.....	306
Tabla 42: Clases de (SNAP) PNI.....	307
Tabla 43: Relaciones de (SNAP) PNI.....	307
Tabla 44: Propiedad tipo dato de (SNAP) PNI.....	307
Tabla 45: Clases de (SNAP) PNS.....	308
Tabla 46: Relaciones de (SNAP) PNS.....	308
Tabla 47: Propiedades tipo dato de (SNAP) PNS.....	308
Tabla 48: Clases de (SNAP) PNT.....	309
Tabla 49: Relaciones de (SNAP) PNT.....	309
Tabla 50: Clases de POWER.....	310
Tabla 51: Relaciones de POWER.....	311
Tabla 52: Propiedades tipo dato de POWER.....	312
Tabla 53: Vocabularios reutilizados en POWER.....	312
Tabla 54: Clases específicas de PROLES.....	313
Tabla 55: Clases de PROLES heredadas de varios vocabularios.....	314
Tabla 56: Relaciones específicas de PROLES.....	315
Tabla 57: Propiedades tipo dato de PROLES heredadas de varios vocabularios.....	317
Tabla 58: Propiedades tipo dato heredadas de TIME INTERVAL.....	318
Tabla 59: Vocabularios reutilizados en PROLES.....	319
Tabla 60: Herramientas reutilizadas por los vocabularios a mapear con ONA y ONAPE.....	320
Tabla 61: Enlaces de interés de los vocabularios a mapear con ONA y ONAPE.....	321

1. INTRODUCCIÓN

One of the most difficult and persistent problems of information can be illustrated by a very simple question. What is the text about? (...) Still, aboutness is at the heart of representing, organizing, and interpreting information, and we must resolve it if we are to retrieve and use information (Raber, 2003, p. 131)

1. INTRODUCCIÓN

La realidad político-económica es, probablemente, una de las más interesantes e influyentes en todo el mundo. Las decisiones de quienes gobiernan, los efectos de sus actuaciones en las personas y en la historia, plagada de guerras, manifestaciones, ataques, etc., son cuestiones que deben ser tratadas y analizadas para saber qué ocurre. En este sentido, la Real Academia Española (RAE) define “política” como: «actividad de quienes rigen o aspiran a regir los asuntos públicos; actividad del ciudadano cuando interviene en los asuntos públicos con su opinión, con su voto, o de cualquier otro modo (RAE, 2014h). Por otro lado, la información económica es el corazón de la información política. La una no se puede analizar sin la otra, y su imbricación tiene unas raíces tan entrecruzadas que cualquier análisis social o político de las noticias de actualidad [...] no sería completo si no se explicaran sus ingredientes económicos» (Río Cabrerizo, 2014, p. 24). La importancia de este dominio, en todos los niveles de la vida, se puede traducir en una necesidad: información organizada y accesible, que facilite la contextualización y comprensión de los acontecimientos.

Entender qué ocurre a nuestro alrededor es posible gracias, entre otros, a los medios de comunicación. El consumo de noticias es una de las formas de conectar con la realidad (Hernández-Serrano, Renés-Arellano, Graham, y Greenhill, 2017). La prensa ejerce como constructora de la realidad social, de informadora de la vida cotidiana, social y política (Bourdieu, 1977, p. 161; Souza Brandão Guaraldo y Almeida Júnior, 2010, p. 198; Troncy, 2008), más aún con las nuevas tecnologías y la multiplicación de información y ruido que dificulta su análisis. Quizá por ello, la política es el dominio más rico en contenidos en la Web (Abadal, Guallar y Codina, 2014; Saleh y Al-Khalifa, 2009). También por su influencia, tanto en la esfera privada como en la pública, se encuentra entre los temas más relevantes y prolíficos en medios de comunicación (Moreira, Batista, Carvalho, Couto y Silva, 2013; Río Cabrerizo, 2014, p. 13), probablemente por el actual entorno cambiante.

Por otro lado, los avances tecnológicos y, particularmente Internet y la Web, han multiplicado la cantidad de información disponible y sus fuentes de acceso. Si bien este problema no es nuevo, pues ya en los años 50 y 60 el crecimiento de la ciencia y su comunicación comenzaron a ser explorados, hoy se produce y comparte información y datos más que nunca (Stuart, 2016, pp. 2-4) y, con ello, aumenta su descontrol y desorganización. La mayor parte de la Web está diseñada por humanos y orientada a la

1. INTRODUCCIÓN

lectura por humanos, lo que ha ido conformando una caótica biblioteca de enormes dimensiones, de la que resulta complicado y costoso extraer conocimiento (Díaz Nosty, 2013; Mannens et al., 2013; Nies et al., 2012; Kallipolitis, Karpis y Karali, 2012; Rodríguez Perojo y Ronda León, 2005; Silva, Souza y Almeida, 2008; Castells et al., 2006; Schwartz, 2001, p. 12). Este problema es planteado por Pastor-Sánchez (2011) de la siguiente forma: «existen cantidades enormes de recursos desorganizados, duplicados o desactualizados, entre los que encontrar la información buscada termina resultando un trabajo arduo y en ocasiones, hasta laberíntico. Rodríguez Mateos y Pérez Lorenzo (2011) califican esta problemática como hiperinflación de contenidos audiovisuales, cuyo exceso provoca que se llegue a desconocer su existencia». En relación con lo anterior, se habla de infoxicación (intoxicación por información) (Cornella, 2010) o sobresaturación de información (*information overload*) (Toffler, 1973, pp. 247-250) para referirse a la incapacidad del ser humano de manejar el volumen de información actual. Es evidente la necesidad de agrupar la información a nivel semántico (Paepen, 2002; Pereira y Baptista, 2003, 2004), esto es, la descripción de los documentos para facilitar su recuperación posterior.

Este contexto no surge con las nuevas tecnologías de la información y la comunicación (TIC), pero sí se acelera por la multiplicación de posibilidades de publicación y distribución de la información. De hecho, algunos de los problemas y necesidades ya fueron resaltados por Litton (1971, p. 42): volumen creciente de documentos, necesidad de acceder a resultados de investigaciones de distinta procedencia, especialización y división del trabajo; necesidad de acceso rápido; necesidad de divulgación rápida, asegurar la autenticidad de lo divulgado y exhaustividad en la recuperación de la información. Litton también destacaba el importante vínculo entre información y valor económico. Las herramientas de representación del conocimiento continúan haciendo frente a cuestiones similares a las planteadas por este autor. Persisten «inconvenientes en aspectos aparentemente sencillos como la recolección, almacenamiento y representación de los datos que se manejan» (Bernal Torres, Castro Romero y González Sanabria, 2017).

Si bien bibliotecas, archivos y otros centros especializados han desarrollado métodos y herramientas para tratar de paliar este problema desde mucho antes de la aparición de Internet, estos no son suficientes para cubrir el volumen de información disponible y desorganizada. Buena parte de los métodos de recuperación de información, fundamentados en algoritmos de localización de términos o cálculo de frecuencias, entre otros, son útiles, pero tampoco aportan soluciones definitivas. La limitación en el acceso más adecuado a los documentos web aún reside en la falta de semántica inteligible por una máquina, lo cual supone un impedimento para la automatización de ciertas tareas. Hay una necesidad de herramientas que ayuden en la recuperación de información, incrementando la precisión sin impactar excesivamente en la exhaustividad (Stuart, 2016, p. 2).

1. INTRODUCCIÓN

Hace unos años se abrió un nuevo campo de experimentación que Berners-Lee, Hendler y Lassila denominaron en 2001 web semántica y definieron después como «una extensión de la actual Web en la que a la información disponible se le da un significado bien definido. Y añaden que su objetivo es que datos definidos y enlazados puedan ser utilizados con mayor efectividad en el descubrimiento, automatización, integración y reutilización por diversas aplicaciones» (Hendler, Berners-Lee y Miller, 2002).

Desde los noventa del pasado siglo, la web semántica ha sido un área puntera de investigación de interés para diversas disciplinas científicas como Documentación, Ingeniería del Conocimiento o Lingüística. Uno de los objetos de investigación en esta área son las ontologías, «una forma de ayudar a controlar algunos de los problemas proporcionando una estructura [para la información, de forma que ésta pueda ser] leída automáticamente y sin ambigüedad, y compartida ampliamente» (Stuart, 2016, p. 4). La creciente popularidad de este tipo de herramientas de gestión del conocimiento se debe, probablemente, a sus enormes posibilidades y a la comprensión compartida de un dominio tanto por personas y como por aplicaciones informáticas, especialmente las ontologías ligeras. Éstas han sido diseñadas para un uso sencillo, accesibles para humanos pero procesables por máquinas, centrándose en las clases principales y propiedades, más que en restricciones y acciones (Rocha, 2012; Stuart, 2016, p. 14). Si «modelar es el proceso de organizar el conocimiento para su uso por una comunidad» (Allemang y Hendler, 2011, p. 24), la ontología es su aplicación práctica. Se ha establecido por sí misma como el elemento más importante de la web semántica (Hlomani, 2014, p. 39), con «un amplio campo de desarrollo» (Rodríguez Mateos y Pérez Lorenzo, 2011). En el ámbito de la Documentación, por sus posibilidades de interoperabilidad y prestaciones, es la herramienta que, con mayor probabilidad, termine imponiéndose, más aún cuando tiene funcionalidades y características propias del resto de instrumentos de representación, combinando «a la vez las características de un lenguaje documental [...] y de un sistema de metadatos» (Codina y Pedraza-Jiménez, 2011).

Es necesario evolucionar hacia una «web donde el mercado de los documentos [en su sentido más amplio] se realice de forma “semántica”, [expresando] el significado de los elementos» (Pedraza-Jiménez, Codina y Rovira, 2007). La web semántica es «una Web de significados, con múltiples fuentes describiendo lo que la información en la Web quiere decir» (Allemang y Hendler, 2011, p. 76). Se habla entonces de la necesidad del modelado de dominios de conocimiento como medio para alcanzar dicha web. Los periódicos y el género informativo por excelencia, la noticia, tienden a expresar información muy compleja en términos actualizados y comprensibles (Codina, 2010), simplificando la narración e interpretación del hecho tratado. De esta forma, la información periodística puede plantearse como fuente para la extracción de términos y conceptos del dominio política-economía y su utilización para la construcción de las ontologías.

1. INTRODUCCIÓN

Estas herramientas de representación del conocimiento contribuyen a la solución de los problemas planteados. El análisis y descripción de acciones político-económicas y su modelado en una ontología de dominio, resulta útil para el control, descripción y la recuperación de información especializada y, con ello, para un mejor conocimiento de la realidad. De hecho, incorporar una ontología en un sistema de información proporciona mejores resultados y más exactos en comparación con métodos tradicionales y aunque haya documentos de texto completo (Codina y Pedraza-Jiménez, 2011; Shoval, Maidel y Shapira, 2008). En esta línea, García-Torres, Pareja-Lora y Pradana-López (2008) destacan que «cada vez son más necesarias ontologías que representen el conocimiento de los distintos dominios que componen la realidad, y que sirvan de base para la anotación de documentos y recursos de la web semántica».

Si bien ya existen ontologías en el dominio política-economía y, sobre todo, en medios de comunicación escrita, muchas se limitan a prototipos, no están disponibles para su acceso y reutilización, se centran únicamente en relaciones jerárquicas o trabajan subdominios concretos. De hecho, «un problema no precisamente menor es que el término tiene tal atractivo (...) que lo utilizan cuando debieran hablar de clasificaciones o de tesauros» (Codina y Pedraza-Jiménez, 2017). Otro tipo de herramientas, como el tesoro EUROVOC o secciones específicas en clasificaciones como la Clasificación Decimal Universal (CDU), están limitadas por las características inherentes a esas herramientas (ausencia de tipos de relaciones, orientadas a un uso humano, etc.). La riqueza de las relaciones e información implica que una ontología es válida no sólo para la indización de recursos, sino que podría ser una base de conocimiento para el descubrimiento de información por sí misma (Stuart, 2016, p. 9). Una ontología, a diferencia de un tesoro, «no está limitada a tres tipos de relaciones [...]. La relación entre dos conceptos puedes ser de cualquier tipo» (Lambe, 2007, p. 238) lo que incrementa las posibilidades de representación y recuperación de información en política y economía. Es destacable, además, su potencial de reutilización, compartición e interoperabilidad (Martínez-Béjar, Ibañez-Cruz, Compton y Cao, 2001).

Detrás de esta tesis no deja de existir también un interés personal en la mejor organización y descripción de documentos del dominio. Ésta es fruto de una preocupación por el actual contexto político/económico de caídas de gobierno, guerras, terrorismo, restricción de derechos sociales, primas de riesgo, etc. Diariamente, la ciudadanía recibe múltiples impactos informativos y las opciones de búsqueda y recuperación de información precisa son insuficientes. Herramientas que faciliten el acceso a la información precisa son necesarias para, como sujetos políticos, tener una visión completa de los acontecimientos. También para conformar una opinión contrastada. Las ontologías, en tanto que vocabularios que permiten representar el conocimiento en un lenguaje formal, parecen ser herramientas cada vez más utilizadas. Por otro lado, el desarrollo de vocabularios controlados es algo en lo que profesionales de bibliotecas, archivos y centros de

1. INTRODUCCIÓN

documentación tienen mucha experiencia y están cada vez más interesados en acercarse a la web semántica (Stuart, 2016, p. 47). Es el momento de ser parte de esta nueva etapa de la gestión de información.

Por todo ello, la investigación se orienta hacia el modelado de acciones del dominio política-economía. Así, se formula la siguiente cuestión ¿cómo se puede modelar una ontología de acciones del dominio política-economía teniendo como base de conocimiento noticias de dicho dominio? En conexión con lo anterior, se plantea si el modelado conceptual que se define garantiza una ontología de acciones en el área de política-economía. Esto es, ¿permite la ontología describir acciones y elementos que las rodean adecuadamente?, ¿permite representar la realidad descrita en las noticias?

1.1. OBJETIVOS

El objetivo general de esta tesis es proponer el modelado conceptual de una ontología para la representación de acciones en el dominio político-económico con el propósito de facilitar la comprensión de los acontecimientos del mundo real. De esta forma, se pretende contribuir al área con una herramienta que dé soporte y mejore la descripción de documentos del dominio. También que facilite la organización de la información y su recuperación en sistemas de información. Este objetivo, se desglosa en los siguientes objetivos específicos:

1. Conocer las metodologías de modelado de ontologías ya existentes para, a partir de las mismas, seguir ciertos pasos para el modelado y construcción de la ontología que se propone. Como uno de estos pasos, se plantea definir un marco teórico de análisis basado en la Gramática de casos orientado a la representación de acciones con ontologías;
2. Conocer ontologías generales, ontologías del dominio política-economía y ontologías de medios de comunicación escrita existentes y valorar posibilidades de reutilización en el modelo planteado;
3. Conocer qué esquemas de metadatos son utilizados en prensa para representar el contenido de las noticias, cuál es su grado de normalización y el potencial de uso de los datos recogidos por los metadatos para evaluar el modelo desarrollado;
4. Validar la adecuación del modelo en un contexto de uso controlado.

1.2. HIPÓTESIS DE TRABAJO

Como hipótesis de trabajo se plantea si, utilizando como fuente de adquisición de conocimiento noticias de medios de comunicación impresa y combinando indización humana y automática, es posible modelar una ontología del dominio política-economía que represente acciones de dicho dominio. Para ello, se aplica una metodología basada fundamentalmente en los principios metodológicos de Mendonça y Almeida (2014): a) revisión teórica relacionada con la web semántica, las ontologías y la indización; b) configuración de una muestra de noticias de periódicos; c) la definición de un marco teórico para el modelado de acciones basado en el análisis de la Gramática de Casos de Fillmore (1967) y García-Gutiérrez (2014); de esta forma, se define, construye e implementa ONA (ONtología de Acciones); d) construcción de una base de conocimiento mediante indización humana y semiautomática, orientada al modelado e implementación de ONAPE (Ontología de Acciones de Política y Economía); e) análisis de vocabularios preexistentes, principalmente de *Linked Open Data* (LOD) para el mapeo inicial de estos vocabularios; f) evaluación basada en datos de ONA y ONAPE. Esta metodología es descrita con mayor detalle en 3.4. DESARROLLO DE LA INVESTIGACIÓN.

1.3. MARCO DE REFERENCIA UNESCO

Este proyecto es multidisciplinar, abarcando varios campos científicos, pues comprende contenidos especializados (Política y Economía), en un contexto digital, con características propias de la noticia (Lingüística, Periodismo), y desde una perspectiva documental y tecnológica (Documentación y, específicamente, Documentación de medios de comunicación e Informática). Conforme a los códigos UNESCO, la tesis propuesta puede enmarcarse en Ciencias Económicas (53), Lingüística (57), Ciencia Política (59), pero principalmente en el campo de las Matemáticas y, concretamente, en Ciencia de los Ordenadores (1203).

1.4. ESTRUCTURA DEL TRABAJO

Se describen brevemente los capítulos de este documento y cómo se organiza esta memoria.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA da comienzo con el análisis de las características de los vocabularios de representación del conocimiento, presentando la ontología como un último estadio de su evolución. También se describe qué son los metadatos, en su vertiente de descripción del contenido de documentos. Después, en el apartado 2.1. LA WEB SEMÁNTICA Y OWL se incide en las características de la web semántica y en el papel del lenguaje OWL en su desarrollado. En el apartado 2.2.

1. INTRODUCCIÓN

ONTOLOGÍAS. CONCEPTO Y CARACTERÍSTICAS se profundiza en el concepto de ontología y en sus características, incluyendo tipología, estructura y metodologías de construcción. La segunda parte de esta fundamentación tiene que ver con otro elemento fundamental en esta memoria, la indización. En este sentido, en 2.3. LA INDIZACIÓN, se define en qué consiste este proceso y se muestran algunas de sus características, distinguiendo entre indización humana y automática. Se incide en la descripción del modelo n-gramas y el software e-Termos, empleado para la extracción de términos en esta investigación. También se describen las particularidades de la indización de noticias y a la Gramática de Casos, empleadas como marco teórico para el modelado de una ontología general de acciones (ONA) y otra de acciones de política-economía (ONAPE). Finalmente, se recogen algunas medidas básicas de evaluación de la indización, que son utilizadas para analizar la capacidad de los vocabularios modelados para describir documentos (en este caso, noticias) del dominio política-economía.

El capítulo 3. PROCEDIMIENTOS METODOLÓGICOS se inicia con el apartado 3.1. ASPECTOS METODOLÓGICOS PREVIOS, referente a las actividades realizadas y conocimientos adquiridos durante la etapa de investigación. A continuación se indica cómo se ha configurado la investigación, esto es aproximación, naturaleza de la investigación, tipo de investigación y procedimientos orientados a la construcción de los modelos planteados. Estos aspectos se detallan en 3.2. TIPO DE INVESTIGACIÓN. En 3.3. LIMITACIONES DE LA INVESTIGACIÓN se exponen algunas de las limitaciones de esta investigación, entre las que se encuentran la imposibilidad de acceso a las noticias a texto completo de algunos periódicos o las dificultades derivadas de la configuración de una muestra multilingüe. Después, se especifican las limitaciones a las que se ha enfrentado esta investigación. En 3.4. DESARROLLO DE LA INVESTIGACIÓN se describe la metodología de modelado e implementación de ONAPE. Es decir, se especifican los pasos que se han seguido hasta la creación de la documentación de los vocabularios modelados. El último apartado de este capítulo, 3.5. APLICACIONES Y PÚBLICO OBJETIVO se dedica a las aplicaciones futuras de ONA y ONAPE como herramienta para la descripción de documentos, entre otros, así como su público objetivo.

Estos pasos y los resultados intermedios y finales se muestran en 4. RESULTADOS. De acuerdo con cada etapa de desarrollo, se definen diversos apartados relativos a la construcción de la muestra, al modelado de ONA basado en el marco teórico de la gramática de casos, al descubrimiento de conocimiento para el modelado y construcción de ONAPE, etc. También se muestran los resultados de la evaluación de los modelos desarrollados. Los componentes de este capítulo se recogen con mayor detalle en 3.4. DESARROLLO DE LA INVESTIGACIÓN.

1. INTRODUCCIÓN

Por lo que respecta al capítulo 5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES, este se estructura en tres apartados. En primer lugar, en el apartado 5.1. CONCLUSIONES se indican las principales conclusiones alcanzadas, una vez concluida la fase investigación. En 5.2. TRABAJOS FUTUROS se lanzan algunas propuestas de investigación relacionadas con la construcción de las ontologías planteadas. Estas propuestas inciden principalmente en la revisión y actualización de ONAPE, pero también se plantean otras cuestiones como su evaluación mediante herramientas especializadas. Por último, en 5.3. PUBLICACIONES CIENTÍFICAS es una exposición de los trabajos presentados durante el periodo doctoral.

En RESUMO EM PORTUGUÊS se presenta un resumen de esta memoria en portugués, como una de las lenguas habituales para la comunicación científica en su campo de conocimiento (Artículo 31. Mención internacional del título de doctor de la [Resolución del Rector de la Universidad de Murcia \(R-310/2015\), de 15 de mayo, por la que se ordena la publicación en el Boletín Oficial de la Región de Murcia de la modificación del Reglamento de Doctorado](#)).

Como ANEXOS se recoge aquellas partes de la investigación que por su extensión se ha considerado adecuado incluir este bloque del documento.

*

**

2. FUNDAMENTACIÓN TEÓRICO- METODOLÓGICA

Existe una tercera hipótesis, realista y concreta, que podría con el tiempo llegar a ser muy factible. En este caso, la mesa de trabajo no se encuentra ocupada por ningún libro. En su lugar existen una pantalla y un teléfono. A lo lejos, en un edificio inmenso, se encontrarían todos los libros y todas las informaciones, con todo el espacio que requiere su registro y su mantenimiento, con todos sus catálogos, bibliografías e índices, con toda la redistribución de datos sobre fichas, hojas y expedientes, con la elección y la combinación realizadas por un personal permanente muy cualificado (Otlet, 1934, p. 429)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Desarrollar estructuras para organizar el conocimiento es un proceso complejo que implica tomar decisiones sobre aspectos semánticos, técnicos y de creación de contenido (Abbas, 2010, p. 44), en un contexto determinado y para unas necesidades concretas. A lo largo del tiempo se han desarrollado diferentes vocabularios para la descripción de documentos, como las Listas de Encabezamientos de Materia (LEM) o la Clasificación Decimal Universal (CDU) y formas más complejas, como los tesauros y ontologías, estas últimas surgidas al calor del desarrollo informático de los años 90. En este capítulo se muestran algunos de los tipos de vocabularios más comunes para la descripción de documentos. Posteriormente, se describe brevemente la web semántica, ámbito en el que las ontologías se desarrollan. A continuación, se dedica un apartado específico para estas herramientas. Finalmente, se analizan conceptos y características de la indización humana y automática, como técnica para la descripción del contenido de los documentos, poniendo el foco de atención en la indización de noticias y en la Gramática de Casos como marco teórico.

Tradicionalmente, en Documentación han existido tres tipos de términos de indización, definidos como «vocablos o expresiones elegidas para representar una noción contenida en un documento» (Fondin, 1977; citado en García Gutiérrez, 1999, p. 111): unitérminos, palabras clave y descriptores.

- **Unitérmino:** vocablo o término simple en lenguaje natural, extraído de un documento, que sirve para representar su contenido. A priori no está normalizado, por lo que puede aparecer en singular, plural, masculino, femenino, verbos (sustantivados o no sustantivados), etc.;
- **Palabra clave:** término, simple o compuesto, asignado a un documento para representar su contenido. Generalmente, se extrae del documento sin una lista previa de la que seleccionar el término, por lo que no existe control. Es lo que (Fondin, 1977) denomina palabras clave libres o derivadas. Sin embargo, el autor también distingue las palabras clave controladas, que se extraen de un vocabulario creado a priori y que implica cierta normalización, al menos en la forma de presentar el término;

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

- **Descriptor:** término, simple o compuesto, asignado a un documento para representar su contenido a partir de un tesoro. Este término es seleccionado de entre un conjunto de sinónimos (no-descriptores) como representante del concepto al que se refiere. Es decir, atiende al principio de univocidad semántica: los conceptos están representados por términos y cada término debería representar a un solo concepto (Alvite Díaz y Martínez González, 2015). El resto de términos sinónimos, no son descriptores, por lo que no pueden ser utilizados para indizar un documento.

Cuando estos términos de indización se agrupan para ser utilizados en procesos de indización se habla de Lenguajes Documentales. Por lo que respecta a su clasificación, Codina y Pedraza-Jiménez (2015, 2017) enmarcan los lenguajes documentales en la semántica documental: «el campo de estudios y de aplicaciones profesionales vinculado con la representación de documentos, tanto con la representación de su contenido (información) como con su identificación y descripción en tanto que objetos (soporte)». Esta rama distingue entre lenguajes documentales, para describir el contenido del documento, y esquemas de metadatos, para identificar el objeto en sí (análisis formal). Con el desarrollo de la web semántica y el auge de las ontologías, estos autores se decantan por el término Lenguajes de Representación del Conocimiento (LRC). Basándose en McGuinness, establecen un *continuum* de los Lenguajes Documentales, situando las listas y taxonomías en un extremo, los tesauros en un punto intermedio y las ontologías en el otro extremo (Figura 1), relacionando estos elementos.

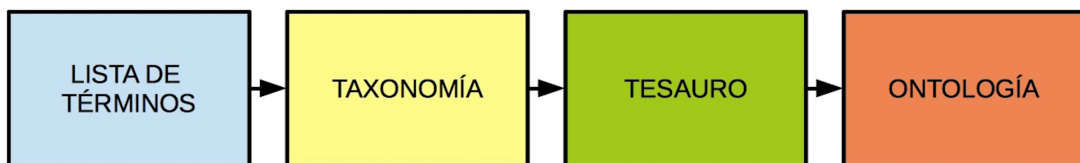


Figura 1: Continuo de los Lenguajes Documentales (Codina y Pedraza-Jiménez, 2015)

No son los únicos autores que han establecido vínculos entre ontologías y vocabularios propios de la Documentación. Para García-Gutiérrez (2004) es una forma más de describir formalmente el conocimiento de un dominio. Currás (2010, p. 57) indica al respecto que existe una «estrecha relación entre ontologías, taxonomías y tesauros: todas ellas tienen una estructura a base de unidades conceptuales». Por su parte, Moyano Grimaldo (2017) concibe las taxonomías como formas primarias de ontologías y Stuart (2016, p. 9) considera ontologías, vocabularios controlados y taxonomías términos cuasi-sinónimos.

Para Codina y Pedraza Jiménez (2015) este Continuo, ya denominado de los Lenguajes de Representación del Conocimiento, incorporan las listas de términos, como una etapa anterior a las taxonomías y considera estas últimas como tipos de clasificaciones.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

A continuación se describen las listas de términos, taxonomías, tesauros y ontologías. A esta descripción se incorporan los esquemas de metadatos, utilizados en la estructuración y descripción de contenidos, especialmente en el ámbito web. El uso de unos u otros dependerá del contexto y necesidades, «una forma más simple de vocabulario controlado, una lista de autoridades o un tesoro puede ser suficiente» (Stuart, 2016, p. 23).

Una lista de términos es un conjunto de términos (simples o compuestos) entre los que no se establece ningún tipo de relación, «muy raramente relaciones de equivalencia entre no-descriptores y descriptores» (Slype, 1991, p. 14), utilizados para representar el contenido de un documento. Se deduce que, generalmente, una lista de términos está compuesta por palabras clave (o unitérminos), sin normalización. Para facilitar su acceso, suelen estar ordenadas alfabéticamente.

Tipos de listas de términos son las listas de unitérminos, las listas de palabras clave y las listas de autoridades, como el catálogo de autoridades de la Biblioteca Nacional de España¹. La Figura 2 muestra la ficha de Cervantes.

Cervantes Saavedra, Miguel de (1547-1616)	
Usado por:	Cervantes (1547-1616) Cervantes, Miguel de (1547-1616) Cervantes di Saavedra, Michele (1547-1616) Сервантес Сааведра, Мигель де (1547-1616) Θερβόντες, Μιγκέλ ντε (1547-1616) Servantes Saavedra, Migel (1547-1616) Therbantes, Minkel nte (1547-1616) Zerbantes eta Saabedra, Mikel (1547-1616) Sirfantis Saafedrā, Mīgīl dī (1547-1616) Sewantisi Saweidela, Migai'er de (1547-1616) سرفنتس ساافدرا، ميغيل دي (1547-1616) 塞万提斯·萨维德拉，米盖尔·德 (1547-1616)
Fuentes:	Don Quijote de la Mancha, 1995; port. (Miguel de Cervantes Saavedra) Los trabajos de Persiles y Segismunda, 1984; port. (Miguel de Cervantes) Persil i Sigismunda, 2008; port. (Migel de Servantes Saavedra) Don Kichote nte la Mantsa, 2009; port. (Μιγκέλ ντε Θερβόντες = Minkel nte Therbantes) Hitroumnyj idal'go Don Kihot Lamančskij, 1997; vol. 1, port. (Miguel' de Servantes Saavedra = Мигель де Сервантес Сааведра) Al-sayyid al-'abqarī Dūn Kījūtih dī lā Manšā, 1957; port. (ميغيل دي سرفنتس ساافدرا = Mīgīl dī Sirfantis Saafedrā) Qī sī yī xiang de shen shi tang Jihede-de-lamanque, 2006; port. (米盖尔·德·塞万提斯·萨维德拉 = Migai'er-de-Sewantisi-Saweidela = Miguel de Cervantes Saavedra) Dic. de literatura española e hispanoamericana, 1993; (Cervantes Saavedra, Miguel de (Alcalá de Henares, Madrid, 1547-Madrid, 1616))
Otro identificador	
normalizado:	http://viaf.org/viaf/17220427 viaf 0000000121221919 isni
Nº Registro:	XX1718747

Figura 2: Ficha de Cervantes en catálogo de Autoridades de la Biblioteca Nacional de España

1 <http://catalogo.bne.es/uhtbin/authoritybrowse.cgi> (último acceso: 07/04/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Otro tipo de lista es la *folksonomía*, «fenómeno que ha tenido lugar en los últimos años» (Yedid, 2016). Se trata de un conjunto de palabras clave asignadas por los internautas de forma colaborativa y compartida para indizar documentos. Parece una adaptación informal a los entornos digitales de la indización con términos (es decir, en lenguaje natural) (Soler Monreal y Gil Leiva, 2010). La idea es que los usuarios generen y clasifiquen los contenidos como quieran. La parte positiva de las *folksonomías* es su grado de receptividad (*openness*) y el hecho de reflejar los conocimientos, vocabulario y necesidades de información de los usuarios. La parte negativa se manifiesta con los problemas de control del lenguaje y su inconsistencia terminológica (Cleveland y Cleveland, 2013, pp. 262-263).

La principal limitación de las listas de términos es la ausencia de relaciones, sólo en algunos casos se contemplan relaciones de equivalencia.

Una taxonomía o clasificación es una lista de términos organizados jerárquicamente. Habitualmente incluye relaciones de clase y de parte-todo. Se describen los tipos de relaciones jerárquicas. Las enumerativas NO son utilizadas en la taxonomía, aunque son incluidas a efectos aclaratorios:

- **Relación jerárquica de clase** (denominada hiperonimia en el ámbito de la semántica lingüística (Pedraza-Jiménez et al., 2007)). Un ejemplo básico es la que se establece entre “sistema de gobierno” y “monarquía parlamentaria”, donde “monarquía parlamentaria” es una clase de “sistema de gobierno”.
- **Relación jerárquica partonómica o de parte-todo** (también llamada meronimia (Pedraza-Jiménez et al., 2007)). Es el caso de la relación que se establece entre “Gobierno” y “ministra”, donde “ministra” esParteDe “Gobierno”.
- **Relación jerárquica enumerativa** (conocida como instanciación o especificación) es la que «identifica la relación existente entre una categoría general de objetos o acontecimientos, expresados mediante un sustantivo común, y un caso individual de tal categoría, que constituye un ejemplo o clase de un solo elemento representado por un nombre propio» (AENOR, 1990, p. 102). Un ejemplo es la relación entre “Mariano Rajoy Brey” y “político”.

En el ámbito periodístico destaca NewsCodes, antes conocida como Topic Sets (Fernández-García, Blázquez, Sánchez-Fernández y Bernardi, 2007). Se trata de un conjunto de vocabularios diseñados por la International Press Telecommunication Council (IPTC), taxonomías utilizadas, entre otros, por Kallipolitis et al. (2012); Fernández-García, Fuentes, Sánchez y Fisteus (2010); Troncy (2008); Kallipolitis, Karpis y Karali (2007); Fernández-García et al. (2006); García, Perdrix y Gil (2006); Schranz, Dustdar y Platzer (2005) y Pereira y Baptista (2003) para la categorización de noticias. NewsCodes está compuesta por términos de diversos dominios, clasificados jerárquicamente en tres niveles,

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

de más general a más específico: *subject*, *subject matter*, *subject detail*. De NewsCodes destacan Subject Reference System (SRS) SubjectCodes y Media Topics, desarrollados a partir del anterior. La Figura 3² muestra el término “sistemas políticos” en SubjectCodes.

Concept ID (QCode) = subj:11024001, ID (URI) = http://cv.iptc.org/newscodes/subjectcode/11024001			
Type: cpnat:abstract	created: 2003-04-05T12:00:00+00:00	modified: 2012-07-10T12:00:00+00:00	retired:
Name (es): sistemas políticos			
Definition (es): Sistema diseñado para brindar orden al gobierno.			
Broader concept: subj:11024000			
Related concept (skos:broader): subj:11024000			
Related concept (skos:exactMatch): medtop:20000653			
Member of scheme: http://cv.iptc.org/newscodes/subjectcode/			

Figura 3: Descripción del término "sistemas políticos" en NewsCodes

Cada término tiene nombre, fecha de creación y de modificación, una breve definición, se indican elementos más genéricos, otros relacionados, etc.

Una evolución interesante es la taxonomía facetada navegacional, a medio camino entre las taxonomías y los tesauros. Maculan (2014, p. 78) la define como:

una estructura de organización compuesta de un sistema de categorías, de forma que, dentro de cada una, se crea una jerarquía de facetas y subfacetas, obedeciendo a criterios preestablecidos para la concepción de las mismas (categorías), o que permite atribuir distintas dimensiones (multidimensionalidad) a un objeto (documento).

Un ejemplo de este tipo de taxonomía es TAFNEGA (Maculan y Lima, 2011), diseñada para facilitar la recuperación de información de tesis y tesis de máster de la Universidade Federal de Minas Gerais (UFMG), en la línea “Organización y uso de la información”.

La principal limitación de las taxonomías, en general, es la ausencia de relaciones, más allá de las relaciones de jerarquía.

Por lo que respecta al tesoro, una de las definiciones más conocidas es la de Slype (1991, p. 10), «lista estructurada de conceptos, destinados a representar de manera unívoca el contenido de los documentos y de las consultas dentro de un sistema documental determinado, y a ayudar al usuario en la indización de los documentos y consultas». Un tesoro no recoge la terminología ni los conceptos del conocimiento universal, sino que captura el de una disciplina, contexto específico o profesión que utiliza un sistema de organización del conocimiento (Abbas, 2010, p. 110). Esta herramienta está compuesta por términos que representan conceptos, relaciones entre estos y notas (de definición, de explicación, históricas). Los tipos de relación que contempla, según Codina y Pedraza-Jiménez (2017), Stuart (2016, p. 8) y Chowdhury (2010, pp. 162-166), son:

2 SubjectCodes: <http://cv.iptc.org/newscodes/subjectcode/> (último acceso: 18/05/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

- **De jerarquía.** Pueden ser relaciones jerárquicas de clase y relaciones jerárquicas partitivas (todo-parte), pero NO enumerativas.
- **De equivalencia.** Se dan entre términos que representan un mismo concepto (sinónimos). Se distingue entre descriptores (términos preferentes), elegidos para representar los conceptos sin ambigüedad (Cavalcanti, 1978, p. 31) y no-descriptores (no-preferentes), que son el resto de sinónimos. Por ejemplo, con País USE Estado, un documento debe indizarse por Estado y no por País. La norma ISO 25964 (AENOR, 2014) tiene en cuenta cuatro grados de equivalencias para analizar los problemas propios del lenguaje natural (García-Marco, 2014): exacta, inexacta (cuasiequivalencia), parcial (genérico-específica) y no equivalencia.
- **Asociativas.** Se establecen entre conceptos cuando no son relaciones ni de jerarquía ni de equivalencia. La Tabla 1 recoge tipos de relaciones asociativas y ejemplos aplicados, de acuerdo con AENOR (1990), Cavalcanti (1978, p. 43), Lancaster (2002, p. 290), NISO (2005) y Slype. (1991, p. 57).

Tipo de relación	Término 1	Término 2
<i>Acciones complementarias</i>	Enseñanza	Aprendizaje
<i>Acción / propiedad</i>	Limpieza ambiental	Polución
<i>Acción / producto</i>	Lagrimo	Lágrima
<i>Antonimia (en el caso de que no se haya establecido entre los elementos una relación de equivalencia)</i>	Inhibición	Excitación
<i>Causalidad (causa / efecto)</i>	Accidente	Víctima
<i>Concomitancia</i>	Síntoma	Enfermedad
<i>Ejecución (proceso / agente)</i>	Control de la temperatura	Termostato
<i>Instrumentación (proceso / instrumento)</i>	Lubricante	Engrase
<i>Localización (acción / lugar de acción)</i>	Enseñanza	Escuela
<i>Materia prima / producto</i>	Harina	Trigo
<i>Objeto o concepto / acción, proceso o disciplina</i>	Insecto	Entomología
<i>Objeto / aplicación</i>	Adhesivo	Unión
<i>Objeto o concepto / origen</i>	Caviar	Mar Caspio
<i>Objeto o concepto / propiedad</i>	Veneno	Toxicidad
<i>Objeto o concepto / unidad de mecanismo o medida</i>	Temperatura	Termómetro
<i>Proceso / contraagente</i>	Fuego	Retardante
<i>Propósito (acción/objetivo)</i>	Encuadernación	Libro
<i>Similitud (en el caso de que no se haya establecido entre los elementos una relación de equivalencia)</i>	Barco	Buque
<i>Sucesión en el espacio o en el tiempo</i>	Planta	Semilla

Tabla 1: Relaciones asociativas entre descriptores de un tesoro

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Un ejemplo de tesoro es EUROVOC, de la Unión Europea. La Figura 4³ muestra el descriptor régimen político. A la izquierda, aparecen las intrarrelaciones con elementos del tesoro y extrarrelaciones con AGROVOC, UNBIS y ECLAS; a la derecha, la traducción del término a otras lenguas de la Unión.

régimen político	
UF <i>sistema político</i>	
04 VIDA POLÍTICA	EQUIVALENTES LINGÜÍSTICOS
MT 0406 marco político	BG политическа система
NT1 bipartidismo	ES régimen político
NT1 cambio de régimen político	CS politický režim
NT1 democracia popular	DA politisk styreform
NT1 dictadura	DE Regierungssystem
NT1 federalismo	ET poliitiline süsteem
NT1 Gobierno en el exilio	EL πολιτεύμα
NT1 Gobierno insurreccional	EN political system
NT1 monarquía parlamentaria	FR régime politique
NT1 monocracia	GA political system (<i>under translation</i>)
NT1 multipartidismo	HR politički sustav
NT1 oligarquía	IT regime politico
NT1 régimen autoritario	LV politiskā sistēma
NT1 régimen de partido único	LT politinė sistema
NT1 régimen militar	HU politikai rendszer
NT1 régimen parlamentario	MT sistema politika
NT1 régimen presidencialista	NL politiek stelsel
NT1 república	PL system polityczny
RT ciencia política [3611]	PT regime político
	RO sistem politic
URI http://eurovoc.europa.eu/3025	SK politický systém
	SL politični sistem
Has Exact Match	FI poliittinen järjestelmä
<i>Sistemas políticos (AGROVOC)</i>	SV politiskt system
<i>Political regimes (ECLAS)</i>	
<i>sistemas políticos (UNBIS)</i>	SR политички систем
	MK политички систем
	SQ sistem politik

Figura 4: Ficha en EUROVOC del descriptor "régimen político"

EUROVOC ha sido actualizado en varias ocasiones, traducido a todos los idiomas de la Unión y adaptado a la norma ISO 25964-1. Es, además, multidisciplinar (aunque con un carácter político europeísta), multilingüe y polijerárquico.

En esta línea, en los últimos años se ha desarrollado el lenguaje SKOS (*Simple Knowledge Organization System*), consolidado como estándar de la web semántica (Alvite Díaz y Martínez González, 2015). Éste que permite la representación de un tesoro en un lenguaje formal, facilitando su utilización en un contexto informático. Otra limitación es la reducción de las relaciones que no sean ni la jerárquica ni de equivalencia a la relación asociativa. Estos dos, formalidad y desarrollo de las relaciones asociativas son precisamente aspectos tratados por las ontologías.

La última herramienta para la organización del conocimiento a describir es la ontología, de la que se puede destacar como característica distintiva «la riqueza de las relaciones entre términos» (Stuart, 2016, p. 9). Es definida por Guarino (1998) como «vocabulario

3 EUROVOC: <http://eurovoc.europa.eu/drupal/?q=es> (último acceso: 15/06/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

específico que se utiliza para describir una realidad, más un conjunto de supuestos explícitos relativos al significado pretendido de las palabras del vocabulario». De esto se desprende que una ontología se compone de: 1) un conjunto de términos (vocabulario); 2) dotados de significados determinados de alguna forma (asunciones); y 3) para describir una realidad. Como los tesauros, son útiles para extraer y recuperar conocimiento específico (Kallipolitis et al., 2012), pero son más complejas y sus objetivos y posibilidades van mucho más allá.

Las ontologías son semejantes a los vocabularios controlados, desarrollados para su utilización en una disciplina, comunidad de práctica o dominio concreto pero difieren en la forma en que son construidas o representadas (Abbas, 2010, p. 163) y en su potencialidad de uso. Con una ontología, una persona no sólo es relacionada con un evento, también puede estar presente en éste, organizarlo, participar en el evento, ser una autoridad en un evento o, posiblemente, instigarlo (Stuart, 2016, p. 9). Una ontología se compone de clases, propiedades, relaciones, individuos, restricciones y axiomas. Estas herramientas son descritas descrita con mayor detalle en 2.2. ONTOLOGÍAS. CONCEPTO Y CARACTERÍSTICAS. En esta tesis, como se indica al inicio de este capítulo, la ontología es concebida como el último producto en la evolución de los lenguajes de representación del conocimiento (Codina y Pedraza-Jiménez, 2015), con evidentes semejanzas con los otros vocabularios de representación (u organización) del conocimiento, en consonancia con autores como Codina y Pedraza-Jiménez (2017), Stuart (2016), Codina y Pedraza-Jiménez (2011), Currás (2010), García-Torres (2008), Gil Leiva (2008) y García Jiménez (2004).

Un ejemplo de ontología es Event, desarrollada por el Centre for Digital Music (Universidad de Londres) para la descripción de eventos. En la Figura 5 se muestran los elementos de dicha herramienta en Protégé, descrito en 2.2.4. HERRAMIENTAS PARA ONTOLOGÍAS: PROTÉGÉ. En la especificación de Event (Raimond y Abdallah, 2007), el mundo no contiene eventos, sino que son patrones de cambio identificados por los agentes, en los que pueden participar activamente agentes, factores pasivos y productos y se localizan en una extensión espacial y temporal.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA



Figura 5: Algunos elementos de la ontología Event en Protégé

Para finalizar este apartado, se describen los esquemas de metadatos, utilizados para la representación de aspectos formales, administrativos y de contenido de documentos, y muy vinculados a la Web. Estos esquemas están compuestos por metadatos y pueden ser entendidos como conjuntos de atributos que representan un recurso (Abbud Grácio y Fadel, 2010; Chowdhury, 2010, p. 62; Hillman, 2005; Kallipolitis et al., 2012; Yaginuma, Pereira y Baptista, 2003b). Más específicamente se trata de «descripciones estructuradas y codificadas que describen características y propiedades de objetos y recursos para facilitar su localización, recuperación, valoración, administración, persistencia e interoperabilidad» (Pastor-Sánchez, 2011, p. 22). Los metadatos son importantes tanto para los proveedores de servicios de información como para sus usuarios, y sirven para un buen número de propósitos (Schwartz, 2001, p. 10), entre los que se encuentran la identificación y descripción de recursos (*descriptive metadata*), su compartición e interoperabilidad y la recuperación de información (*information retrieval*) (Chowdhury, 2010, pp. 63-71).

El vínculo entre vocabularios y esquemas de metadatos es el marcado semántico de documentos. Los primeros, son usados en los segundos para ese proceso de anotación semántica. En la relación entre ontologías y esquemas, los metadatos únicamente llegan a la descripción de los recursos web en forma de propiedades y los consiguientes valores que toman para cada recurso. Para realizar procesos más complejos, según Pastor-Sánchez

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

(2011, p. 28), es necesario la ontología, un concepto complementario al de metadatos, que introduce la idea de conocimiento estructurado y permite realizar deducciones a partir de axiomas. Esto permite diseñar estructuras complejas de información que, además de facilitar el intercambio y explotación de datos, apunta a la ejecución de inferencias para obtener información que no se ha declara explícitamente. Para Codina y Pedraza-Jiménez (2011) la semántica documental es la combinación de lenguajes de representación del conocimiento y esquemas de metadatos.

La forma en que se asignan propiedades a los recursos con esquemas de metadatos se basa en una idea similar a las sentencias RDF, en las que los recursos son descritos a partir de tripletas sujeto-predicado-objeto (Pastor-Sánchez, 2011, p. 56), como muestra la Figura 6.

<i>Sujeto</i>	<i>Predicado</i>	<i>Objeto</i>
Noticia	Elemento de metadato (propiedad)	Valor que toma el elemento
Noticia	Keywords	Keyword1, Keyword2, etc.

↓

<i>Sujeto</i>	<i>Esquema metadatos</i>	<i>Elemento metadato</i>	<i>Valor</i>
Noticia	p.e. Dublin Core	Keywords	Keyword1, Keyword2, etc.

Figura 6: Descripción de una noticia con metadatos

Cuando se agrupa un conjunto de metadatos para un fin concreto, por ejemplo, para representar un tipo específico de documento, hablamos de esquemas de metadatos. Es decir, preacuerdos que aportan información sobre un recurso tipo mediante la definición de una serie de propiedades específicas. Mediante estructuras jerárquicas de metadatos es posible la representación de información en varios grados de detalle (Jokela, Turpeinen, Kurki, Savia y Sulonen, 2001). Estas no indican «qué sistema debe ser usado para describir el asunto de un documento, [sino que] si se quiere decir algo sobre una materia, se debe hacer en éste o en aquel campo» (Keyser, 2012, p. 164). Se muestran, así, como un elemento intermedio que busca sistematizar y ahondar en las posibilidades de descripción.

Un ejemplo de esquema de metadatos es Dublin Core (DC)⁴, un conjunto de metadatos para la descripción general de recursos en línea, del que también se habla en A. DESCRIPCIÓN DE VOCABULARIOS MAPEADOS. En la Figura 7 se muestra el metadato título de este esquema. En DC, para cada metadato se indica nombre, URI, etiqueta, definición y un comentario que suele aclarar la definición.

4 <http://dublincore.org/documents/dces/> (último acceso: 19/05/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Nombre del término: título	
URI	http://purl.org/dc/elements/1.1/title
Etiqueta	título
Definición	Nombre dado al recurso
Comentario	Habitualmente, un Título será el nombre por el que un recurso es formalmente conocido

Figura 7: Metadato título en el esquema Dublin Core

Los esquemas de metadatos suelen ser utilizados en los códigos fuente de las páginas web. Para ello, es necesario especificar su espacio de nombre (*namespace*). Se trata de «un contenedor abstracto que congrega distintos identificadores únicos, [cada uno con] su propia descripción» y organizados de una forma estructurada (Bernal Torres et al., 2017). Un elemento existe en un espacio de nombre particular (del que se indica su URI) y debe ser validado contra éste. Así se identifica de dónde procede cada elemento y se evitan colisiones por el uso de nombres de elementos comunes (Alesso y Smith, 2004, p. 70; 200). Los espacios de nombre son utilizados no sólo para identificar esquemas de metadatos, también se emplean para identificar elementos de ontologías.

Una limitación de los esquemas es su inflexibilidad, pues su capacidad para describir el contenido de documentos se reduce a las propiedades definidas en el esquema (Pan, 2009, p. 73). Pese a ello, el uso de esquemas de metadatos está bastante extendido. Algunos son analizados en 4.5.1. PREPROCESAMIENTO DE LA SUBMUESTRA2 PARA LA EVALUACIÓN DE ONAPE: METADADOSHTML.

Una variante del esquema es el perfil de aplicación, cuyos componentes proceden de otros espacios de nombre, combinados conjuntamente para un uso particular (Stuart, 2016, p. 15). Un ejemplo es el utilizado en Omnipaper, descrito en 4.4.1. PRESELECCIÓN DE VOCABULARIOS.

El siguiente apartado se centra en la web semántica, marco de desarrollo de las ontologías web y el lenguaje OWL.

2.1. LA WEB SEMÁNTICA Y OWL

La ontología es una herramienta desarrollada fundamentalmente en el seno de la web semántica, «una extensión de la Web actual donde la información viene dotada de significado bien definido y permite a ordenadores y personas trabajar en cooperación» (Berners-Lee, Hendler y Lassila, 2001). El artículo del que proviene esta cita, uno de los precursores de la web semántica, remarcaba que ésta: 1) es una extensión de la web actual; 2) tendría una tendencia hacia la descentralización para facilitar su desarrollo; 3) tiene como propósito el procesado automático de datos e información; 2) aporta estructura al

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

contenido significativo de las páginas web; 3) tiene como reto aportar un lenguaje que exprese tanto datos como reglas para razonar sobre los datos y que permita que reglas procedentes de cualquier sistema de representación del conocimiento sean exportadas a la web; 4) precisa de estructuración y de reglas de inferencia para el procesado automático.

Para hacer esto posible, Berners-Lee (2000) diseña un esquema en el que describe qué tecnologías básicas componen la web semántica, organizadas en capas. Dicho esquema es desarrollado posteriormente por Nowak (2009) y adaptado al castellano por Pastor-Sánchez (2011, p. 36)⁵ en la Figura 8. En Pastor-Sánchez (2011, pp. 34-40) se describe cada componente de la arquitectura de la web semántica:

- RDF es un modelo de datos para identificar y describir recursos;
- RDFS permite describir las propiedades y clases de los recursos;
- OWL aporta un modelo para la representación de ontologías, concebidas como conjuntos de colecciones de informaciones relacionadas entre sí y que contribuyen al significado bien definido de dichas informaciones (Berners-Lee et al., 2001).

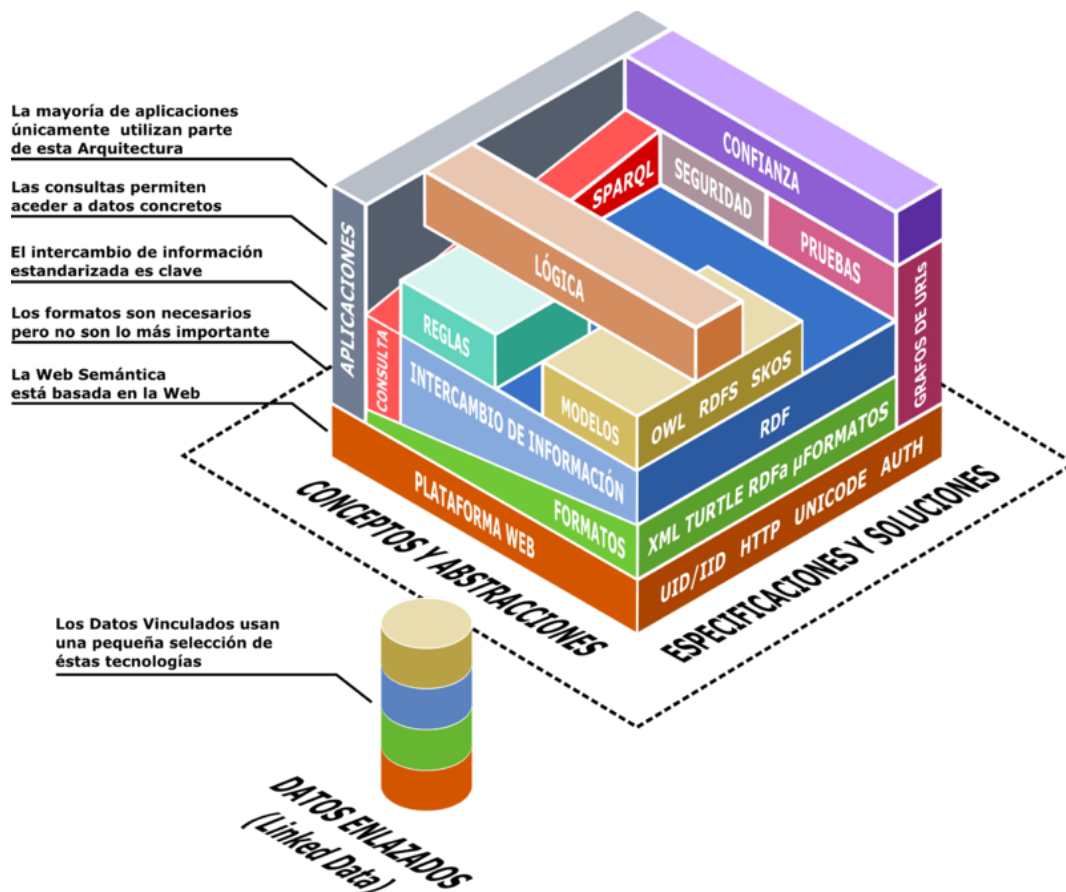


Figura 8: Arquitectura de la web semántica

5 Accesible en <https://commons.wikimedia.org/w/index.php?curid=18602139> (último acceso: 19/05/2016)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Las elementos de la Figura 8 son agrupados por Peis Redondo, Herrera-Vieda, Hassan-Montero y Herrera (2003a) en tres tipos de capas, que se detallan en la Tabla 2.

CAPAS DE DESCRIPCIÓN, ESTRUCTURACIÓN E IDENTIFICACIÓN DE LOS RECURSOS WEB	
1	<p>UNICODE permite la codificación de caracteres en ordenadores con el propósito de asignar un único número y nombre a cada carácter de cada lengua; es independiente del ordenador y de la plataforma utilizados (Pastor-Sánchez, 2011, p. 36).</p> <p>URI (<i>Uniform Resource Identifier</i>) / IRI (<i>Internationalized Resource Identifier</i>) son cadenas de caracteres estandarizadas para identificar recursos de una red de manera unívoca. Si además la URI permite localizarlos, se habla de URL (<i>Uniform Resource Locator</i>). Otro tipo de URI es la URN (<i>Uniform Resource Name</i>), que permite identificar recursos, pero no implica su disponibilidad (Pastor-Sánchez, 2011, p. 37).</p>
2	<p>XML (<i>eXtensible Markup Language</i>) es la base sintáctica sobre la que se sustenta el resto de capas [de la web semántica], es un metalenguaje que permite definir nuestros propios lenguajes de etiquetado (Peis Redondo et al., 2003a). Es un lenguaje de marcas para la estructuración, intercambio y publicación de datos, que permite separar estructura, contenido y aspecto visual de un documento siguiendo un esquema jerárquico en el que unas etiquetas deben estar correctamente incluidas en otras. «Este lenguaje constituye la base para el desarrollo de lenguajes orientados a la representación de información altamente estructurada y con diferentes niveles de abstracción» (Pastor-Sánchez, 2011, p. 54), pero no indica qué significa esta estructura. Por tanto, XML tiene una capacidad limitada para expresar la semántica, pues no diferencia objetos de relaciones ni tiene noción de jerarquía de clases (Berners-Lee et al., 2001; Pastor-Sánchez, 2011, pp. 46-54; Ruíz-Martínez, 2011, p. 19). Los estándares abiertos XMLS y NS amplían sus posibilidades: XMLS (XMLschema) «funciona de manera similar a un esquema de bases de datos en el que el contenido de los elementos tiene asociado un tipo»; NS (<i>Namespaces</i>) son especificaciones que describen cómo asociar una URI con cada etiqueta y atributo en un documento XML</p>
3	<p>RDF (<i>Resource Description Framework</i>) es un lenguaje de marcas en XML, una recomendación de W3C que define un modelo de datos para la descripción estructurada de recursos de información en Internet. Usa: 1) Sentencias o tripletas, compuestas por sujeto, predicado y objeto; 2) Espacios de nombre (NS), mediante los que identifica los elementos de los vocabularios utilizados en las sentencias (Pastor-Sánchez, 2011, p. 56); y 3) URI, que identifican sujetos, propiedades y objetos.</p> <p>RDFs (RDF schema) es un vocabulario RDF que «permite describir recursos mediante una orientación a objetos similar a la de muchos lenguajes de programación» (Peis Redondo et al., 2003a). Proporciona opciones básicas para el modelado, aportando una mayor semántica al definir clases (<i>rdfs:Resource</i>; <i>rdfs:Property</i>; <i>rdfs:Class</i>), propiedades (<i>rdfs:type</i>; <i>rdfs:subClassOf</i>) y restricciones de propiedad. Un esquema permite validar el documento, es decir, si los elementos que incluye se utilizan correctamente.</p> <p>Peis Redondo et al. (2003a) añaden aquí DCMI (<i>Dublin Core Metadata Initiative</i>), que, según Pan (2009, p. 74) aportan alguna semántica informal, pero es inflexible. Las ontologías, según este autor, son una alternativa para especificar el significado de los recursos web.</p>
CAPAS DE DOTACIÓN DE LÓGICA A LA WEB	
4	<p>OWL (<i>Web Ontology Language</i>): es el modelo recomendado para la representación de ontologías (Stevens et al., 2007), una extensión de RDFS con una semántica más formal. Las declaraciones en OWL definen clases, propiedades e individuos (instancias) junto a las propiedades ya definidas en RDFS, y añaden otras como la relación inversa entre propiedades y un conjunto de axiomas adicionales para definir restricciones, como la cardinalidad entre las instancias de clase. También incorpora enumeraciones de instancias de clases y combinaciones de clase (intersección, unión, complemento). Los agentes inteligentes, productos software que recogen, filtran y procesan</p>

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

	información, realizan inferencias e interactúan con su entorno de forma autónoma, requieren de lenguajes ontológicos y ontologías para su funcionamiento (Peis Redondo et al., 2003a).
5	La capa lógica se basa en la capacidad de aplicar reglas de inferencia para inferir nuevo conocimiento a partir de información explicitada. «En el contexto de la web semántica, inferencia quiere decir simplemente que, declarada cierta información, se puede determinar» otra información relacionada que también se puede considerar declarada (Allemang y Hendler, 2011, p. 80). Se considera que las capas ontológica y lógica están integradas, ya que son las ontologías las que permiten definir axiomas lógicos (Ruíz-Martínez, 2011, pp. 20-21).
CAPAS DE DOTACIÓN DE FIABILIDAD, CONFIANZA Y SEGURIDAD DE ACCESO A LOS RECURSOS WEB	
6	La capa de prueba define una estructura en la que un agente es capaz de justificar con pruebas que las inferencias obtenidas en una búsqueda son correctas. Precisa de las reglas de inferencia de la capa lógica, capacidad de rastreo y prueba del origen de una secuencia lógica y de la firma digital, bloques cifrados de información que se podrán usar para verificar que la información adjunta proviene de una fuente fiable (Berners-Lee et al., 2001; Morales-del-Castillo, 2008, pp. 49-52)
7	En la capa confianza se establecen redes sociales de confianza a través de las cuales los agentes determinan el grado de confiabilidad de una fuente. Cualquier usuario, humano o agente, está identificado mediante una URI (Ruíz-Martínez, 2011, p. 21; Morales-del-Castillo, 2008, pp. 49-52).

Tabla 2: Descripción de las capas de la arquitectura de la web semántica

OWL (*Web Ontology Language*), en la capa 4, es el lenguaje más extendido para el modelado y construcción de ontologías por ser, entre otros, la recomendación de la *World Wide Web Consortium* (W3C) para expresar ontologías para la web semántica (Stuart, 2016, p. 15; Hitzler y Parsia, 2009, p. 111; Stevens et al., 2007). Como se ha indicado, OWL descansa sobre XML, XMLS, RDF y RDFS. La aportación de RDF se centra en las opciones de representación de conocimiento mediante tripletas sujeto-predicado-objeto (Allemang y Hendler, 2011, p. 31):

- **Sujeto:** recurso que se describe (persona, lugar, cosa). Puede ser cualquier cosa, identificada unívocamente mediante una URI;
- **Predicado:** propiedad de un sujeto (nombre, ciudad, título, color, forma, característica), identificada unívocamente mediante una URI;
- **Objeto:** valor que toma la propiedad del sujeto (Peter, San José, “Guerra y Paz”, azul, círculo, fuerte). Puede ser cualquier tipo de dato RDF, que soporta cualquier tipo de dato XML (Alesso y Smith, 2004, p. 191).

La W3C indica los conceptos básicos de RDF en (Schreiber y Raimond, 2014), mostrando ejemplos de tripletas RDF: en la Figura 9 *Bob* es un Sujeto o Recurso, *is interested in* es un predicado o propiedad, y *The Mona Lisa* es un Objeto o valor. Para evitar colisiones por el uso de la misma terminología, en la identificación de cualquier elemento de una tripleta en un documento web se utilizan los espacios de nombre (*namespaces*). Estos, definidos en la página 23.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

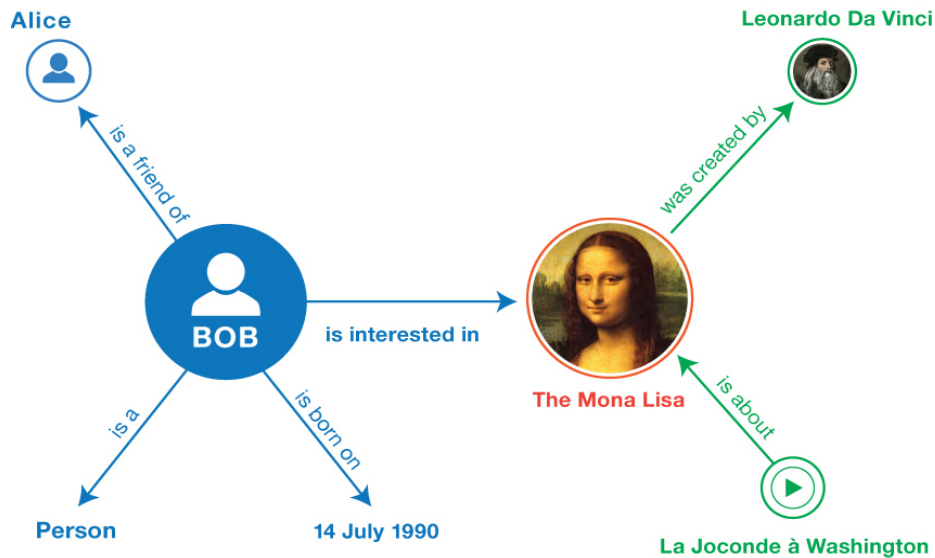


Figura 9: Gráfico informal de tripletas ejemplo

Las opciones de modelado con RDF son básicas, limitadas a la identificación de objetos, relaciones binarias, reificación, contenedores y colecciones. La reificación es un mecanismo de RDF que convierte una sentencia o tripleta en un recurso. Esto es, «hacer declaraciones sobre otras declaraciones» (Peis Redondo, Herrera-Viedma, Hassan-Montero y Herrera, 2003b).

RDFS va más allá y define un modelo que permite crear vocabularios que admiten clases, propiedades y restricciones (Allemang y Hendler, 2011, pp. 127-133; Pastor-Sánchez, 2011, pp. 56-75):

- **Clases** (*rdfs:Class*): representan todos los recursos que tienen unas características comunes. Existen tres clases en RDFS: recurso (*rdfs:Resource*), para cualquier elemento susceptible de ser descrito; propiedad (*rdfs:Property*), clase a partir de la cual se definen propiedades para describir los recursos; clase (*rdfs:Class*), clase utilizada para definir otras clases o instancias.
- **Propiedades** (*rdfs:Property*): representan características de las clases y se definen de manera independiente a éstas. En RDFS se pueden definir tres propiedades: tipo (*rdfs:type*) para relaciones de enumeración; subclase de (*rdfs:subClassOf*) para relaciones de jerarquía de clase; subpropiedad de (*rdfs:subPropertyOf*) para relaciones de jerarquía entre propiedades.
- **Restricciones de propiedad** (*rdfs:constrainProperty*): indican algún tipo de limitación en las propiedades, a través de dos tipos de restricciones: dominio (*rdfs:domain*), que limita las clases a las que se aplican una propiedad; rango (*rdfs:range*), que indica los valores que puede tomar esta propiedad.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Con RDF y RDFS ya es posible crear ontologías simples, e incluso algunas tareas de razonamiento automáticas para inferir nuevas relaciones a partir de una base de conocimiento dada (Ochoa Hernández, 2011, p. 41). Pero cuentan con limitaciones, como la imposibilidad de definir restricciones de rango sólo para algunas clases, no permiten definir clases disjuntas ni combinar clases, tampoco indicar restricciones de cardinalidad o características particulares de las propiedades (Antoniou y Harmelen, 2009, pp. 92-93). Así, para mayores desarrollos se utiliza OWL, «diseñado para representar conocimiento complejo sobre cosas, grupos de cosas y relaciones entre cosas» (Hitzler, Krötzsch, Parsia, Patel-Schneider y Rudolph, 2009).

OWL chocaba con la dificultad de combinar el poder expresivo de RDFS y la lógica total (*full logic*), lo que dio lugar a tres sublenguajes OWL. La elección de uno u otro dependerá de las necesidades que debe atender la ontología: 1) OWL Full: combina las primitivas del lenguaje OWL de forma arbitraria con RDF y RDFS. «Está dirigido a usuarios que desean una expresividad máxima y la libertad sintáctica ofrecida por RDF, aunque sin garantías de cómputo»; 2) OWL DL (*OWL Description Logic*): es un sublenguaje de OWL Full que restringe la forma en que los constructos de OWL pueden ser utilizados en RDF, perdiendo parte de la compatibilidad con este lenguaje; 3) OWL Lite: es una versión aún más limitada, que admite una jerarquía de clasificación y restricciones simples (la restricción de cardinalidad se limita a valores de 0 y 1). Excluye, por ejemplo, las clases enumeradas y las clases disjuntas, con la consecuente pérdida de expresividad (Ruíz-Martínez, 2011, pp. 31-33; Antoniou y Harmelen, 2009, pp. 93-94).

OWL2, una revisión de OWL1, mejora el conjunto de tipos de datos incorporados y las posibilidades de los tres sublenguajes indicados. Esto se traduce en tres perfiles, cuya funcionalidad se orienta esencialmente a ciertos tipos de aplicaciones. Todos ellos «comparten algunas limitaciones en cuanto a su expresividad»: no permiten la negación ni la disyunción (Ochoa Hernández, 2011, p. 52). OWL2-EL es útil para aplicaciones que utilizan ontologías con un abundante número de propiedades y/o clases, refleja el perfil de la familia EL++, perteneciente a la Lógica Descriptiva; OWL2-QL (*OWL2 Query Language*) es adecuado «para aplicaciones que utilizan gran número de instancias de datos y en las que la consulta es la tarea de razonamiento más importante». Es una variante de OWL-Lite; el razonamiento de OWL2-RL (*OWL2 Rule Language*) se puede implementar mediante un lenguaje de reglas estándar. Resulta especialmente adecuado para aplicaciones con ontologías relativamente ligeras y se puede utilizar para organizar un elevado volumen de individuos mediante tripletes RDF (Ochoa Hernández, 2011, p. 52; Ruíz-Martínez, 2011, pp. 31-32).

A continuación se profundiza en el concepto, características, componentes, tipología y otras cuestiones relativas a las ontologías.

2.2. ONTOLOGÍAS. CONCEPTO Y CARACTERÍSTICAS

El término ontología comenzó a ser utilizado en el siglo XVII, en el ámbito filosófico. Fue popularizado por Christian Wolff, ya entrado el siglo XVIII. Como objeto de la Metafísica, rama que estudia la naturaleza, estructura, componentes y principios fundamentales de la realidad, la ontología «se enfrenta a problemas relacionados con la existencia o la naturaleza de las cosas» (Martínez-Béjar, Benjamins y Martín-Rubio, 1997). Esto es, se ocupa de la existencia de los entes y de cómo estos «pueden ser clasificados dentro de una jerarquía y subdivididos de acuerdo a similitudes y diferencias» (Ochoa Hernández, 2011, pp. 3-5).

Teniendo en cuenta lo anterior, se puede entender por qué el término acabó siendo utilizado en el campo de la Inteligencia Artificial en los años 80 del siglo XX (Stuart, 2016, p. 9). Habiéndose extendido el uso del término a otras ciencias, especialmente a la Inteligencia Artificial y la Documentación, todavía sigue generando controversias. Probablemente porque definiciones propuestas tienen una vertiente específica y otra genérica, es decir, por la distinción entre ontologías formales y ligeras. La discusión también tiene lugar entre quienes consideran que herramientas como los tesauros y taxonomías son tipos de ontologías y quienes las ven como productos diferentes (Sánchez-Cuadrado, Morato-Lara, Palacios-Madrid, Llorens-Morillo y Moreiro-González, 2012). El carácter interdisciplinar de estas herramientas da lugar a una heterogeneidad de perspectivas.

Entre las definiciones más extendidas de ontología se pueden destacar las siguientes:

1. Neches et al. (1991) mencionan por primera vez el término ontología en el área de la Inteligencia Artificial (IA) (Ochoa Hernández, 2011, p. 5) y la describen como «un instrumento que define los términos básicos y relaciones que constituyen el vocabulario de un área de conocimiento, así como las reglas de combinación de estos términos y relaciones para definir extensiones al vocabulario».
2. Para Gruber (1993), cuya definición es todavía hoy la más extendida (Stuart, 2016, p. 9), una ontología es «una especificación explícita de una conceptualización». Studer, Benjamins y Fensel (1998) añaden que dicha especificación es formal. Veamos una explicación de esta definición más detallada. Según Pedraza-Jiménez et al. (2007): una conceptualización es «un modelo abstracto de algún fenómeno del mundo construido mediante la identificación de los conceptos relevantes a ese fenómeno (normalmente un dominio del conocimiento)»; explícita significa que los conceptos utilizados en la ontología, y las restricciones para su uso, están claramente definidos; formal indica que debe (...) estar expresada mediante una sintaxis que permita a un ordenador operar sobre ella; compartida expresa que contendrá conocimiento consensuado en algún grado.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

3. Para Uschold y Gruninger (1996) el término ontología se refiere a «un entendimiento compartido de algún dominio de interés que sería utilizado como un marco unificado para resolver problemas», tales como 1) el abordaje de diferentes perspectivas y el uso de distintos términos para describir la misma idea; 2) la conversión de unos formatos de modelos a otros; y 3) la falta de interoperabilidad entre sistemas de base de conocimiento, aportando soluciones como: a) el desarrollo de un marco conceptual unificado, que facilite que los resultados de la investigación en un campo sean aplicados en otros; b) automatizar el proceso de conversión del modelo cuando un software externo sea introducido; y c) utilizar un aproximación basado en agentes federados (*federated agent-based approach*) para la compartición de conocimiento. Para los autores, una ontología encarna necesariamente una conceptualización, esto es, una cierta visión del mundo con respecto a un dominio dado, concebida como un conjunto de conceptos (entidades, atributos, procesos), sus definiciones y sus interrelaciones. Esta ontología puede ser implícita (existe sólo en la cabeza de una persona o está incorporada en un software) o explícita.
4. Desde un punto de vista más formal, Guarino (1998) define ontología como un «vocabulario específico que se utiliza para describir una realidad, más un conjunto de supuestos explícitos relativos al significado pretendido de las palabras del vocabulario». Se desprende que una ontología se compone de un conjunto de términos (vocabulario) dotados de significados, determinados de alguna forma (asunciones), para describir una realidad.

De estas definiciones se puede deducir que una ontología es una representación de cierta realidad, expresada a través de términos; que consta de varios componentes, siendo, al menos uno de estos un vocabulario, que representa una serie de conceptos de cierto dominio, y un conjunto de relaciones entre estos, que son explicitadas; que se expresa en un lenguaje tal que permite su procesado por máquinas; y que se construye a partir de un conocimiento compartido entre personas del dominio de interés.

Se remite, de esta forma, al triángulo semiótico de Ogden y Richards (1923, p. 11) (ver Figura 10), que añaden el *referente* al signo lingüístico de Saussure (1916), compuesto por *significante* y *significado*.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

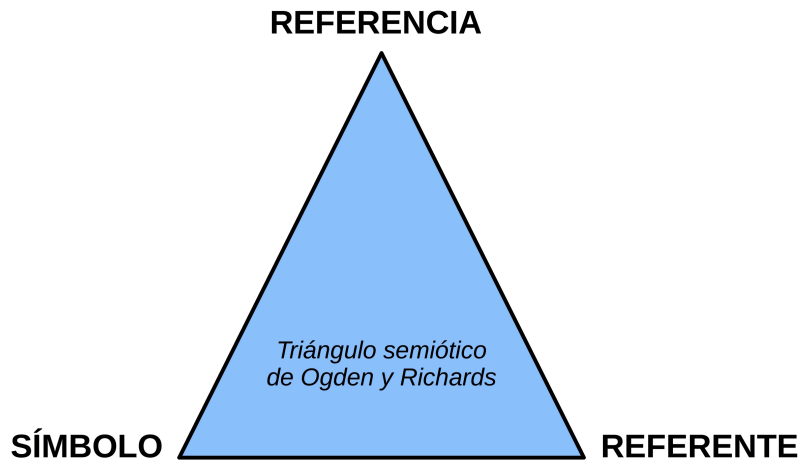


Figura 10: Triángulo semiótico de Ogden y Richards (1923)

En este triángulo, el símbolo (o significante) es cada término que compone el vocabulario; la referencia es el pensamiento o significado determinado por el ser humano, los conceptos; el referente es el objeto (del mundo real o no) al que se refiere el símbolo y que evoca una imagen mental en el individuo. Los tesauros suelen operar en el lado símbolo (términos) – referencia (conceptos) y las ontologías en el de referencia – referente, ya que «tratan de capturar y representar el significado de un dominio del conocimiento en particular y simularlo explícitamente, siguiendo modelos mentales mediante el razonamiento inferencial» (Rodríguez Perojo y Ronda León, 2005).

En el siguiente apartado se incide en la estructura de una ontología y sus elementos.

2.2.1. COMPONENTES DE LA ONTOLOGÍA

La descripción de los componentes de la ontología se basa fundamentalmente en Horridge (2011), así como Allemang y Hendler (2011, pp. 153-185), Gil Leiva (2008, pp. 227-228), Gómez, Moreno, Pazos y Sierra-Alonso (2000), Gruber (1995) y Ochoa Hernández (2011). El trabajo de Horridge es un manual de Protégé, de ahí las menciones al software y los ejemplos utilizados. Una ontología puede tener clases, propiedades, relaciones, individuos, restricciones y axiomas:

1. Clases (*classes*): son *sets* que contienen individuos. Se trata de representaciones concretas de conceptos, que se describen mediante declaraciones que indican los requerimientos para ser miembro de éstas. Protégé contempla los siguientes tipos:
 - **Clases disjuntas** (*disjoint classes*): aquellas de las que un individuo no puede ser instancia simultáneamente. Por ejemplo, las clases *Brazo* y *Pierna* son disjuntas, ya que cualquier miembro que pertenezca a la clase *Brazo* no puede formar parte del conjunto de la clase *Pierna*.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

- **Clases complementarias** (*complement classes*): cuando se define que la clase A es complementaria de la clase B, ningún individuo de la clase A puede ser miembro de la clase B. Por ejemplo, una instancia de la clase *Perro* no puede serlo de la clase *Gato*.

Existen otros tipos de clases, que son definidas a partir de la definición de restricciones (*restrictions*). Dichos elementos describen el conjunto de individuos de dicha clase basándose en las relaciones en las que estos participan:

- **Clases primitivas** (*primitive classes*): hacen referencia a las condiciones necesarias que tienen que reunir los individuos para formar parte de dicha clase. Horridge (2011) recurre al ejemplo de *ChessyPizza*, que tiene 2 condiciones necesarias: 1) *esUna Pizza* y 2) *tieneTopping algún Queso*.
 - **Clases equivalentes** (*equivalent classes*): el simple hecho de cumplir unas condiciones no implica obligatoriamente que un individuo sea miembro de cierta clase. Para ello, estas condiciones necesarias deben definirse como suficientes. En Protégé, una clase que tiene al menos un set de condiciones necesarias y suficientes también se denomina **clase definida** (*defined classes*) (Horridge, 2011, p. 57).
 - **Clases enumeradas** (*enumerated classes*): se crean definiendo previamente el conjunto de individuos (instancias) que forman parte de dicha clase. Por ejemplo *DiasDeLaSemana*, tiene 7 individuos: {*Lunes, Martes, Miércoles, Jueves, Viernes, Sábado, Domingo*}.
2. **Propiedades** (*properties*): también llamadas *slots* o atributos, expresan relaciones entre individuos o entre individuos y datos. Además, se pueden establecer subpropiedades. Las propiedades se definen habitualmente para conjuntos de individuos (esto es, de clase) pero también para individuos concretos (de instancia). Una propiedad tiene un dominio (*domain*) y un rango (*range*). El dominio incluye el conjunto de individuos que se relaciona con el conjunto de individuos contemplados en el rango.

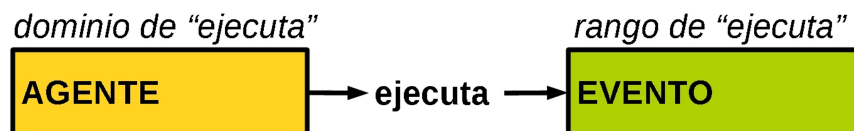


Figura 11: Dominio y rango de la relación o propiedad tipo objeto "ejecuta"

Por ejemplo, en la propiedad *ejecuta* el dominio es el conjunto de individuos de la clase *Agente* y el rango el grupo de individuos de la clase *Evento* (ver Figura 11).

Existen tres tipos de propiedades, de anotación, de tipo dato y de tipo objeto:

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

- **Propiedades de anotación** (*annotation properties*): utilizadas para añadir información a clases, individuos y propiedades de objeto o dato. En esta memoria se utiliza la propiedad *comment* para las descripciones de elementos e *isDefinedBy* para indicar la fuente de esa descripción.
- **Propiedades tipo dato** (*datatype restrictions*): vinculan individuos y valores (datos) definidos mediante un rdf literal o con un esquema de metadatos XML. Para establecer estas relaciones se utilizan las restricciones *tieneValor* (*hasValue restrictions*).
- **Propiedades tipo objeto** (*object type restrictions*): son relaciones binarias entre individuos, interconexiones entre los conceptos en un universo de discurso, normalmente definidas a nivel de clase. Se describen en el siguiente punto.

3. **Relaciones** (*relations*) o propiedades tipo objeto: pueden definirse distintas relaciones entre clases de una ontología. En la Tabla 3, basada en Gómez et al. (2000) y Ochoa Hernández (2011, pp. 11-26), se muestran algunos tipos de relaciones y qué propiedades se suelen definir para éstas.

Relaciones	Propiedades						
	Simetría	Asimetría	Reflexividad	Irreflexividad	Transitividad	Intransitividad	Dependencia
Causal: describe cómo, dados unos estados o acciones, se induce a otros		xx		x		x	x
Condiciona: define las condiciones en las que algo tiene lugar		x		x	x		x
Cronológica: describe la secuencia temporal en la que ocurren las acciones		x		x	x		
Dependencia: tipo especial de relación de asociación a través de responsabilidades, parentesco, propiedades, etc.		x	x		x		
Equivalencia: relación de similitud entre expresiones aparentemente diferentes	x		x		x		
Funcionales: describe las condiciones para las acciones y reacciones que se producen y sus posibles consecuencias		x		x	x		x
Partonómica, de mereología o estructural: relación de jerarquía entre un todo y sus partes		x			x		
Propósito: establece el porqué y el para qué de los conceptos		x		x	x		
Similitud: establece qué conceptos son análogos			x		x		
Taxonómica: relación de jerarquía de clases entre conceptos en el que primero es clase y el segundo subclase		x		x	x		
Topológica: describe una distribución espacial de conceptos físicos e interconexiones entre estos conceptos	x		x		x		

Tabla 3: Tipos de relaciones entre clases y algunas de sus propiedades

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Algunas de las relaciones anteriores son similares a las definidas en un tesoro, ya sean relaciones jerárquicas, de equivalencia o como tipos específicos de relaciones asociativas (ver Tabla 1). Por otro lado, la Tabla 3 muestra algunas de las propiedades de las relaciones de una ontología. En la siguiente descripción se han añadido las propiedades inversa, funcional e inversa funcional:

- **Propiedad simétrica** (*symmetric property*): permite establecer relaciones de igualdad entre dos individuos. Si *María esHermanaDe Luisa* y dicha propiedad es simétrica, *Luisa esHermanaDe María*;
- **Propiedad asimétrica** (*asymmetric property*): hace referencia al caso contrario, si A se relaciona con B mediante una relación asimétrica no puede aplicarse la misma relación entre B y A;
- **Propiedad reflexiva** (*reflexive property*): se aplica en los casos en que un individuo se relaciona consigo mismo. Por ejemplo, una *Institución* puede investigarse a sí misma;
- **Propiedad irreflexiva** (*irreflexive property*): es la propiedad opuesta a la anterior. Por ejemplo, *María esHermanaDe María* es irreflexiva (e imposible);
- **Propiedad transitiva** (*transitive property*): si un individuo A se relaciona con el individuo B y éste con un individuo C, A y C también están relacionados;
- **Propiedad intransitiva** (*intransitive property*): muestra el caso contrario. Si A se relaciona con B y éste con un individuo C, A y C no están relacionados;
- **Propiedad inversa** (*inverse property*): si una propiedad tipo objeto relaciona el individuo A con el individuo B, su inversa relacionará B con A. Por ejemplo, la propiedad *esParteDe* tiene como inversa *tieneComoParte*;
- **Propiedad funcional** (*functional property*) o de valor único (*single valued property*): un individuo se relaciona como máximo con otro individuo, es decir, la cardinalidad de la propiedad es 0 ó 1. *Pizza* siempre *tieneComoBase* una *Base*, y sólo una;
- **Propiedad inversa funcional** (*inverse functional property*): implica que si una propiedad es funcional su inversa también lo es. Un ejemplo clásico es la relación que se establece entre *Pasaporte* y *Persona*, donde *Pasaporte* tiene una relación 1-1 con *Persona* y ésta tiene o no tiene relación con *Pasaporte*. Por el contrario, en *Pizza tieneComoBase Base*, la relación es funcional pero su inversa no, ya que *Base esParteDe* una o más *Pizza*.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

El conjunto de relaciones de una conceptualización se denomina conjunto relacional básico.

Un tipo especial de relación es la función, donde el valor del último concepto es único para el concepto precedente. Ejemplos de funciones son *madreDe* y *precioCocheUsado* (Gómez et al., 2000).

4. **Individuos** o **instancias** (*individuals*): representan objetos concretos de un dominio y están relacionados por enumeración con una clase, de la que dependen jerárquicamente.
5. Las **restricciones** (*restrictions*) determinan qué individuos forman parte de una clase en función de las relaciones en las que éstos participan. Pueden ser de cuantificación, de cardinalidad o *tieneValor*.

Dentro de las restricciones de cuantificación (*quantifier restrictions*) se contemplan las restricciones existenciales y las universales:

- **Restricciones existenciales** (*existential restrictions*), también denominadas *some restrictions* o, en OWL, *someValuesFrom*. Indican que los individuos del dominio de una propiedad se relacionan con al menos un individuo del rango de dicha propiedad. No es excluyente, los individuos del dominio pueden relacionarse con individuos de otras clases. Por ejemplo, un individuo *Universitario* forzosamente *estudiaEn* *some Universidad* pero también *estudiaEn* un *Instituto*. Esta restricción es impositiva, ya que debe haber como mínimo una relación.
- **Restricciones universales** (*universal restrictions*) son denominadas en OWL *allValuesFrom*. Mediante este tipo de restricción, los individuos del dominio de la relación sólo se relacionan con los individuos del rango, excluyendo cualquier relación con otras clases. Si en el ejemplo anterior se define la propiedad *estudiaEn* con una restricción universal, un *Universitario* sólo pueden relacionarse con *Universidad*. Esta restricción no es impositiva, puede haber 0 ó más relaciones. Pero, si existen, serán sólo con los individuos del rango especificado.

Con las restricciones de cardinalidad (*cardinality restrictions*) se limitan el número de relaciones entre individuos. Existen 3 tipos: al menos (*at least*), como máximo (*at most*), exactamente (*exactly*). Por ejemplo, podemos restringir a 2 la relación entre *Tronco* y *Brazo*, donde *Tronco tiene at most 2 Brazo*.

Las restricciones *tieneValor* (*hasValue restrictions*) permiten concretar matemáticamente las relaciones entre clases y valores. Por ejemplo, un *A4* debe

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

cumplir la restricción de medir 297 x 210 mm. Para ello, se establece que *A4* tiene *altura* y *anchura*, limitadas a 297 y 210 mm, respectivamente: *A4 hasAlturaValue some integer [=297]* y *A4 hasAnchuraValue some integer [=210]*.

6. Los axiomas son expresiones que siempre son verdaderas y pueden ser implementadas en las ontologías de distintas formas. Por ejemplo, Protégé permite generar particiones de valor (*value partitions*) (Horridge, 2011, pp. 67-70), que pueden refinar las descripciones de clases. Dentro de las mismas, los axiomas de cobertura (*covering axioms*) son clases concebidas como la unión de dos clases, generando una superclase respecto de las clases que cubren. De esta forma, un axioma de cobertura se compone de dos elementos: la clase que está siendo cubierta y las clases que conforman la cobertura.

Los componentes de una ontología puede variar dependiendo del tipo y características de ésta. Así, hay quienes distinguen la estructura de una ontología y las instancias. Por ejemplo, una ontología de libros podría no contener ninguna información sobre libros concretos, pero sí la estructura necesaria para describirlos, así como las relaciones entre estos y tipos asociados de objetos. Bajo esta perspectiva, se puede hablar de un conjunto de elementos de ontología (*ontology element set*) y de las instancias de ontología (*ontology instances*). La combinación de ambos proporciona un conjunto de datos de ontología (*ontology data set*) o base de conocimiento (*knowledge base*) (Stuart, 2016, pp. 12-21).

Ésta es una diferencia fundamental entre tesauros y ontologías, junto con la especificidad e las relaciones entre clases y la posibilidad de establecer inferencias. La inferencia es la capacidad de un sistema para definir nuevas relaciones en un conjunto de datos a partir de las relaciones existentes y un conjunto de reglas. Por ejemplo, si una ontología sobre libros cuenta con los elementos «Charles Dickens ha escrito *Tiempos difíciles*» y la relación *ha escrito* sólo es utilizada para expresar relaciones entre autores y trabajos, se puede inferir que Charles Dickens es un escritor y “Tiempos difíciles” un trabajo (Stuart, 2016, p. 19).

Por lo que respecta cómo denotar clases y propiedades en la identificación de un elemento de la ontología en OWL, éstas no pueden tener espacios ni caracteres especiales (Nicolino y Ferneda, 2014). A este respecto, se han desarrollado diferentes estilos de citación. Horridge (2011, pp. 17, 26) recomienda la notación *CamelBack*, con dos convenciones (no obligatorias): los nombres de clases deben comenzar con una letra mayúscula y sin espacios, por ejemplo, *PizzaTopping*; para las propiedades, la recomendación es comenzar el nombre de la propiedad en minúscula, sin espacios e iniciar las siguientes palabras con mayúscula, por ejemplo, *isIngredientOf*. Para describir estos elementos, «OWL cuenta con la propiedad [opcional] label [etiqueta en español], con la que es posible la definición exacta del término que identifica una entidad (clase, individuo o propiedad), permitiendo el uso del espacio y de letras acentuadas» (Nicolino y Ferneda, 2014).

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Para estas etiquetas, Aguado-de-Cea, Montiel-Ponsoda, Poveda-Villalón y Giraldo-Pasmin (2015) recomiendan:

- Utilizar *labels* de clases tan cortas como sea posible, autocontenidas, significativas y concisas, capturando o resumiendo de forma concisa el significado de la clase;
- No emplear abreviaturas, ya que el principal objetivo de las *labels* es describir y aportar significado a la entidad, para garantizar un uso apropiado y la adopción de la ontología por usuarios/as finales;
- Incluir varias *labels* (sinónimos o variaciones del término) si son utilizadas como equivalentes en cierto dominio-tareas;
- Las etiquetas de clases deben estar en singular y, preferentemente, aceptar sufijos que faciliten la derivación de los términos: *creator, creation, is creator of, etc.*;
- Las *labels* de las propiedades deben formar una unidad sintagmática compuesta por, al menos, un verbo y la unidad sintagmática que lo acompaña, normalmente un objeto que representa el argumento más cercano al verbo y/o preposición. Por ejemplo: *tiene* o *tiene sede en*;
- Las *labels* deben reflejar, en los idiomas basados en el género, dicho género. Por ejemplo: *está ubicada en, está ubicado en*.

El siguiente apartado está dedicado a las tipologías de ontologías.

2.2.2. TIPOS DE ONTOLOGÍAS

En la literatura se muestra un buen número de clasificaciones de ontologías, en función de diversos criterios. Las distintas tipologías pueden superponerse o utilizar denominaciones distintas para la misma idea. A continuación se muestran algunas de las clasificaciones existentes:

Por lo que respecta al grado de desarrollo, Benjamins y Gómez-Pérez (2000) distinguen entre ontologías ligeras (*lightweight*) y ontologías pesadas (*heavyweight*). Las primeras, de uso más extendido especialmente en la web semántica (Stuart, 2016, p. 14), definen un vocabulario y algunas especificaciones sobre los significados de los términos que recoge. Las segundas incluyen axiomas y mecanismos de inferencia y tienen un mayor grado de formalización.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Según Guarino (1998) existen cuatro tipos de ontologías, según el nivel de generalización:

- **De alto nivel:** describen conceptos generales, tales como espacio, tiempo, materia, objeto, hecho, acción, etc., son independientes de un dominio en concreto;
- **De dominio:** se circunscriben a un dominio de conocimiento concreto, mediante la especialización de términos definidos en ontologías de alto nivel;
- **De tareas:** describen una tarea o actividad genérica, por ejemplo, ventas, también mediante la especialización de términos definidos en ontologías de alto nivel;
- **De aplicaciones:** describen conceptos dependientes tanto del dominio como de la tarea y, con frecuencia, son especializaciones de ambos tipos de ontologías. Habitualmente estos conceptos se corresponden con el papel que ciertas entidades de dominio desempeñan al ejecutar una tarea.

Benjamins y Gómez-Pérez (2000) refieren un listado con los tipos de ontologías en función de su grado de reusabilidad:

- **Ontologías de representación del conocimiento:** capturan las primitivas de representación utilizadas para formalizar conocimiento en paradigmas de representación del conocimiento;
- **Ontologías generales:** incluyen un vocabulario sobre cosas, eventos, tiempo, espacio, causalidad, comportamiento, funciones, etc.;
- **Ontologías de alto nivel:** proporcionan nociones generales bajo las que todos los términos de ontologías deben estar vinculadas;
- **Meta-ontologías,** también llamadas ontologías genéricas o fundamentales: son reutilizables a través de dominios. La ontología de mereología sería el ejemplo típico;
- **Ontologías de dominio:** reutilizables en un dominio dado, proporcionan vocabularios sobre los conceptos de un dominio y sus relaciones, sobre las actividades que tienen lugar en dicho dominio y sobre las teorías y principios elementales que gobiernan dicho dominios;
- **Ontologías de tareas:** proveen de un vocabulario sistemático de términos utilizado para resolver problemas asociados a una tarea particular, dependiente o no de un dominio. Están orientadas a la resolución de problemas;
- **Ontologías de dominio-tareas:** son ontologías de tareas reutilizables y un dominio dado, pero no en cualquier dominio;

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

- **Ontologías de método:** proporcionan definiciones de conceptos relevantes y relaciones utilizados por un método para lograr su propósito. Están orientadas a la resolución de problemas;
- **Ontologías de aplicación:** contienen el conocimiento necesario para modelar una aplicación en particular.

Esta clasificación se basa en la idea de reutilización de ontologías para la construcción de nuevos vocabularios. Las ontologías de representación del conocimiento serían las que aportarían el mayor grado de reusabilidad y las de aplicación, las que menos, pues atenderían a propósitos muy concretos.

Roussey, Pinet, Kang y Corcho (2011, pp. 10-25) distinguen cuatro formas de definir un concepto en una ontología: mediante definiciones textuales, un conjunto de propiedades, una definición lógica (compuesta por varias fórmulas) o un grupo de instancias que pertenecen a la clase. Los autores denominan intensionales las tres primeras definiciones de un concepto y extensional la última definición. Ello, unido a las relaciones entre conceptos y/o instancias, determina los elementos que componen una ontología y su nivel de formalización:

- **Ontologías de información:** se componen de diagramas y esbozos (*sketches*) usados para clarificar y organizar las ideas de quienes intervienen en un proyecto, del que se muestra una vista general con conceptos, instancias y sus relaciones;
- **Ontologías terminológicas / lingüísticas:** como glosarios, diccionarios vocabularios controlados, taxonomías, *folksonomías*, tesauros y bases de datos léxicas. Estas ontologías presentan y definen el vocabulario que utilizan, cuya terminología es el resultado de un acuerdo entre usuarios de una comunidad, con el fin de evitar la ambigüedad terminológica;
- **Ontologías de software o dirigidas a la implementación de software:** proporcionan esquemas conceptuales con la idea de almacenar y manipular datos en actividades de desarrollo de software, para asegurar la consistencia de los datos. Aquí los conceptos están compuestos por conjuntos de propiedades y son definidos mediante relaciones entre ellos. Estas relaciones tienen restricciones;
- **Ontologías formales:** requieren una semántica inequívoca para usar un lenguaje con el fin de definir conceptos, motivaciones claras para diferenciar entre conceptos adoptadas, así como reglas estrictas sobre cómo se refinan conceptos y relaciones.

Para modelar y construir ontologías es preciso seguir una metodología, una guía que oriente y sistematice qué pasos seguir en el proceso. El siguiente apartado presenta brevemente algunas de estas metodologías.

2.2.3. METODOLOGÍAS PARA CONSTRUIR Y EVALUAR ONTOLOGÍAS

Existen múltiples acercamientos en el modelado y construcción de ontologías, que pueden diferir, por ejemplo, en función del punto de inicio. En este sentido, García-Torres et al. (2008) indican cuatro procedimientos: desde cero, reingeniería de ontologías (reutilizando otras preexistentes), métodos basados en el aprendizaje automático o aquellos basados en la fusión de ontologías. Así, por ejemplo, Beydoun, López-Lorca, García-Sánchez y Martínez-Béjar (2011) plantean una metodología de construcción de ontologías jerárquicas (taxonomías) desde cero, a partir de un conjunto de conceptos definidos por sus atributos, sobre los que aplican una serie de algoritmos para el análisis de similitudes entre estos. Esta metodología, como señalan los autores, puede ser utilizada para mejorar ontologías ya existentes.

Sin embargo, en la creación de ontologías también se pueden combinar dos o más procedimientos, no existe ninguna metodología definitiva, aunque sí varias propuestas con un uso más o menos extendido (Silva, 2008, pp. 120-121). Diversos trabajos apuntan las causas de esta falta de estandarización, entre las que se indican la falta madurez de las aportaciones (Benjamins y Gómez-Pérez, 2000), carencias en los procedimientos descritos (Silva et al., 2008) o la aplicación a proyectos concretos, que dificultan su reutilización (Iqbal, Murad, Mustapha, Sharef y Mohd, 2013).

En realidad, el modelado y construcción de una ontología es más un proceso artesanal que una actividad de ingeniería (Benjamins y Gómez-Pérez, 2000), donde el análisis del dominio y la evaluación de la ontología son dos de los pasos que requieren mayor esfuerzo (Simperl, Mochol y Bürger, 2010). En este sentido, para Keyser (2012, p. 134) «las ontologías podrían ser extraídas automáticamente de los textos o, incluso, complementadas de forma automática pero, por ahora, son escritas por humanos».

En la literatura pueden encontrarse varias publicaciones que analizan estas metodologías. Uno de los trabajos más exhaustivos es Silva (2008, pp. 218-263), que identifica y describe hasta nueve metodologías y desarrolla su propia propuesta a partir de las anteriores. También Iqbal et al. (2013) y Mendonça y Almeida (2014) se detienen en el análisis de trabajos previos. A modo de ejemplo, en la Tabla 4 se presentan cuatro de las metodologías más referenciadas: TOVE (Grüninger y Fox, 1995), ENTERPRISE (Uschold y Grüninger, 1996; Uschold y King, 1995), Methontology (Gómez-Pérez, Fernández y Vicente, 1996) y Método 101 (Noy y McGuinness, 2001). Puede encontrarse información más detallada sobre éstas y otras en la bibliografía citada y en trabajos como Haav (2011), entre otros.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Metodología	Características	Etapas
<i>TOVE</i>	<p>Origen: proyecto <i>Toronto Virtual Enterprise</i> (TOVE)</p> <p>Objetivo: crear un modelo de <i>senso comune</i> compartido y corporativo</p> <p>Dominio: negocios (empresarial)</p>	<ol style="list-style-type: none"> 1. Elaboración de escenarios de motivación 2. Cuestiones de competencia de información 3. Concepción formal de la terminología 4. Cuestiones de competencia formales 5. Especificación de axiomas formales 6. Verificación de teoremas completos
<i>ENTERPRISE</i>	<p>Origen: <i>Enterprise Ontology</i></p> <p>Objetivo: desarrollar la ontología indicada, de conocimientos del ámbito corporativo</p> <p>Dominio: negocios (empresarial)</p>	<ol style="list-style-type: none"> 1. Propósito, grado de formalismo y usuario 2. Construcción de la ontología: conceptualización; implementación, integración con ontologías preexistentes 3. Evaluación 4. Documentación
<i>Methontology</i>	<p>Origen: desarrollo de la metodología</p> <p>Objetivo: construcción de ontologías mediante reingeniería de ontologías, utilizando conocimiento de dominio</p> <p>Dominio: diversos</p>	<ol style="list-style-type: none"> 1. Especificación 2. Adquisición de conocimiento 3. Conceptualización 4. Integración 5. Implementación 6. Evaluación 7. Documentación
<i>Método 101</i>	<p>Origen: desarrollo de ejemplos prácticos utilizando Protégé</p> <p>Objetivo: construcción de ontologías de dominio</p> <p>Dominio: diversos</p>	<ol style="list-style-type: none"> 1. Determinación del campo de aplicación 2. Consideración de la reutilización de términos 3. Enumeración los términos 4. Definición clases 5. Organización de las clases en una taxonomía 6. Definición de propiedades y sus restricciones 7. Adición de valores a las instancias

Tabla 4: Ejemplos de metodologías de construcción de ontologías

Tanto Silva (2008) como Silva et al. (2008) y Mendonça y Almeida (2014) destacan Methontology como la metodología más completa. No obstante ninguna está totalmente desarrollada. Dado que falta un estándar, parece más práctico recurrir a unos principios metodológicos que a una metodología concreta.

Estos principios, definidos por Mendonça y Almeida (2014) permiten seguir una serie de fases en el modelado y construcción de una ontología, especificando las características generales de la ontología; las técnicas y fuentes que van a ser utilizadas para la adquisición de conocimiento; qué términos forman parte de la ontología y qué relaciones se establecen entre estos; cómo se formalizan los elementos de la ontología; qué lenguaje formal se utiliza para implementar la ontología y cómo se va a evaluar. El último paso hace referencia a la documentación de la ontología. Estos principios se resumen en la Tabla 5.

Stuart (2016, pp. 100-116) destaca el descubrimiento de conocimiento como un tipo de adquisición de conocimiento (fase 2 en la Tabla 5), que consiste en la extracción de

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

información de fuentes existentes. Asimismo, indica que, para esta cuestión, el texto puede ser analizado manualmente, semiautomáticamente o automáticamente. Pero si el volumen de documentos a estudiar es relativamente pequeño, puede ser más rápido el análisis manual que instalar y probar la amplia variedad de soluciones informáticas disponibles.

Por lo que respecta a la formalización de la ontología, indica Stuart que será necesario definir, a partir de los términos obtenidos en la fase anterior, nombres, etiquetas y otros elementos, como las definiciones, que ayudarán a darle consistencia. Además de propiedades y clases, se suelen utilizar las *rdfs:label* y *rdfs:comment*, dominio, rango y cardinalidad de las propiedades (Stuart, 2016, pp. 117-119).

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Fase	Descripción
Fase 1 <i>Especificación de la ontología</i>	<ul style="list-style-type: none"> - Propósito general de la ontología - Objetivos específicos - Tipo de ontología y alcance - Escenarios de uso - Tipos de usuarios de la ontología - Grado de formalidad
Fase 2 <i>Adquisición de conocimiento</i>	<ul style="list-style-type: none"> - Adopción de métodos de adquisición del conocimiento: análisis de textos del dominio, entrevistas con especialistas, grupos focales con expertos, cuestionarios, análisis y técnicas de inducción, tormenta de ideas - Selección apropiada de materiales de referencia para la obtención del conocimiento propio del dominio
Fase 3 <i>Fundamentación de la ontología</i>	<ul style="list-style-type: none"> - Selección de una o más perspectivas filosóficas para el desarrollo de la ontología - Evaluación del uso y aplicación de una o más ontologías de fundamentación en el desarrollo de la ontología <p>Esta fase puede ser cuestionada</p>
Fase 4 <i>Conceptualización de la ontología</i>	<ul style="list-style-type: none"> - Identificación de los términos del dominio para la ontologías a partir de los materiales de referencia utilizados. Implica investigar estos términos (o sinónimos) en otras ontologías para su reutilización y, si no existen, crear una clase nueva - Construcción de una taxonomía de clases del dominio (relación <i>isA</i>) - Construcción de otras relaciones como la paronimia (relación <i>todoParte</i>) y estructurales generales de representación (<i>produce</i>, <i>participaEn</i>, etc.)
Fase 5 <i>Formalización de la ontología</i>	<ul style="list-style-type: none"> - Definición formal de las propiedades de clases y relaciones. Los autores sugieren: identificador de la clase (ID), nombre de la clase (label), ontología de origen (<i>importadoDe</i>), URI (<i>hasURI</i>), sinónimos (<i>hasSynonim</i>), término alternativo, definición textual, aclaración, ejemplo de uso, definición formal, anotaciones - Definición formal de las clases y relaciones: a partir de la definición textual de cada clase y relación se construye una definición formal que mediante relaciones ontológicas sustituye los verbos y relaciona los términos de la definición - Construcción de axiomas formales - Definición de las instancias para cada clase
Fase 6 <i>Implementación de la ontología</i>	<ul style="list-style-type: none"> - Selección de un lenguaje formal - Generación de la representación formal de la ontología en el lenguaje escogido, utilizando un editor de ontologías
Fase 7 <i>Evaluación de la ontología</i>	<ul style="list-style-type: none"> - Análisis de inconsistencias, incompletitudes y ambigüedades de los términos de la ontología mediante los razonadores de los editores de ontologías - Especificación de cuestiones de competencia de validación de la ontologías - Utilización de criterios de evaluación de la calidad de la ontologías - Consulta con expertos del dominio para la validación del contenido de la ontología, mostrándoles un interfaz visual de la misma
Fase 8 <i>Desarrollo de la documentación de la ontología</i>	<ul style="list-style-type: none"> - Elementos de especificación - Cuestiones de competencia para la definición del alcance y validación - Descripción de ontologías reutilizadas - Relación de términos reutilizados - Diccionarios de clases - Diccionario de relaciones, jerarquías y otras estructuras de representación - Criterios, métodos y técnicas utilizados en la evaluación y validación - Relación y descripción de los materiales de referencia para la adquisición de conocimiento

Tabla 5: Principios de construcción de ontologías (Mendonça y Almeida, 2014)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

La evaluación de una ontología implica dos partes: verificación y validación (Gómez-Pérez, 2004), citada en Hlomani (2014, p. 22) y Vrandečić (2009, p. 295). La verificación de una ontología tiene que ver con su corrección, esto es, si satisface los requerimientos definidos en los estadios iniciales de su modelado. Para esta cuestión, editores de ontologías como Protégé incluyen razonadores que asisten la evaluación desde esta perspectiva, como HermiT. También existen herramientas específicas, como OOPS (*Ontology Pitfall Scanner!*)⁶ donde, con sólo introducir la URL de la ontología o directamente pegar el código en el apartado correspondiente, analiza (verifica) si existe algún error en la ontología de entre los cuarenta y uno tipos implementados.

Por el contrario, la validación analiza si la ontología es la correcta, es decir, si realmente modela el dominio para la que fue diseñada, pues ésta puede estar influida por quien participa en su construcción. El modelado de este tipo de vocabularios conlleva, además, cierta carga de subjetividad derivada de las experiencias del/a ontólogo/a, así como de sus conocimientos previos, tanto tecnológicos como del dominio. Por otro lado, el contexto en el que se modela, desde un punto de vista temporal, espacial y cultural también cambia, pues el propio dominio evoluciona. Por ello, es importante evaluar la adecuación de cualquier herramienta de representación del conocimiento al dominio en que se circunscribe, analizar si es útil para esa función de representación de los documentos. En este sentido, Hlomani (2014) desarrolla un marco de validación de ontologías basado en datos, comparando una ontología de dominio con datos sobre dicho dominio.

De esta forma, el autor concibe la evaluación de ontologías como un método que analiza «el grado o más bien la distancia entre esta conceptualización aproximada (el modelo) y el mundo real» (Hlomani, 2014, p. 52). La utilidad de este método reside en la posibilidad de combinar metodología de construcción de la ontología y validación, pues esta última tiene en cuenta aspectos esenciales en la primera: los *corpora* y los métodos de análisis de textos para la extracción de conceptos. Para ello, se basa en una serie de elementos entre los que destaca la cobertura como la medida más relevante y propone como métricas los sesgos temporal y categorial (ver Tabla 6).

Métrica	Medidas
<i>Exactitud (accuracy)</i>	Precisión, Exhaustividad, Cobertura
<i>Compleitud (completeness)</i>	Cobertura
<i>Concisión (conciseness)</i>	---
<i>Consistencia (consistence)</i>	Recuento: Número de términos con significados inconsistentes
<i>Sesgo temporal (temporal bias)</i>	Cobertura
<i>Sesgo categórico (category bias)</i>	Cobertura

Tabla 6: Medidas para evaluar ontologías desde el punto de vista de su corrección

6 <http://oops.linkeddata.es/> (último acceso: 19/05/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

A este respecto, ONA y ONAPE son validadas mediante exhaustividad y precisión, que miden la exactitud de la ontología, según Hlomani (ver 2.3.4. EVALUACIÓN DE LA INDIZACIÓN). También se emplea la medida-F, que combina las anteriores.

El método de Hlomani primero evalúa cómo varias ontologías cubren el dominio de flujo de trabajo (*workflow domain*) en determinados intervalos de tiempo o categorías; después analiza cómo la cobertura de cada ontología cambia en el tiempo o a través de las subcategorías del dominio. La cuestión temporal es importante en cualquier dominio y tipo documental. Por ejemplo, (dos Santos y Bräscher, 2015) destaca la necesidad de evaluar, en el contexto de los conjuntos de documentos periodísticos, las variaciones y cambios lingüísticos que tienen lugar con el paso del tiempo. Los datos, basados en la similitud entre ciertos términos seleccionados en las ontologías y en la muestra, evidenciaron importancia estadística en la diferencia entre las coberturas de las ontologías a través de los intervalos de tiempo y con respecto a las subcategorías de un dominio cuando las ontologías son comparadas entre sí. No halló, sin embargo, estadística significativa cuando se analiza la cobertura de una ontología a través de las subcategorías del dominio.

El autor concluye que en la evaluación basada en datos se deben tener en cuenta las dimensiones del dominio, tanto en el predesarrollo de una ontología, su desarrollo y posteriormente. Más allá de las métricas destaca, al modelar y construir una ontología, la importancia de: 1) Seleccionar una fuente válida para la extracción de los términos; 2) Seleccionar como documentos de los que extraer términos aquellos que tengan elementos en común. Además, demuestra que las ontologías y el dominio cambian con el tiempo y en función de las subcategorías. Por ello, la ontología responderá mejor al momento actual si se seleccionan documentos del mismo intervalo temporal y de fuentes del mismo tipo de información. Estos aspectos se traducen en la necesidad de describir fuentes y documentos de los que se extraen términos para construir la ontología. En esta tesis se tienen en cuenta estas conclusiones en 4.1. CONFIGURACIÓN DE LA MUESTRA, donde se muestra cómo se ha construido la muestra de noticias y se recogen las características de éstas como fuente de términos para el modelado de ontologías. Como limitación, destaca la subjetividad, presente no sólo en el modelado, también en la validación de la ontología en el momento en que se seleccionan criterios, límites en los criterios y en sus resultados.

A continuación se describe someramente Protégé, un editor de ontologías, utilizado habitualmente en la implantación de ontologías, esto es, en la fase 6 de la Tabla 5.

2.2.4. HERRAMIENTAS PARA ONTOLOGÍAS: PROTÉGÉ

Existen varias herramientas para la construcción de ontologías (García Jiménez, 2004), dependiendo de los lenguajes de formalización, intereses y necesidades. Así, existen productos con licencia comercial y otros libres. Unos programas obligan a utilizar su propio lenguaje mientras que otros son independientes y se basan en un lenguaje

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

estandarizado. García-Torres et al. (2008) señalan varias opciones disponibles, como, por ejemplo, Protégé, WebOnto o KAON. Protégé es el software más extendido en la fase de adquisición del conocimiento (en la fusión y alineamiento de ontologías) y en la construcción de ontologías de dominio. Probablemente, porque es gratuito y de código abierto (Keyser, 2012, pp. 134-135). Además, que permita editar ontologías en distintos lenguajes ontológicos y sus opciones de personalización facilitan que los/as usuarios/as puedan modificar la ontología en otros editores dependientes del lenguaje (Mizoguchi y Kozaki, 2009, pp. 328-329).

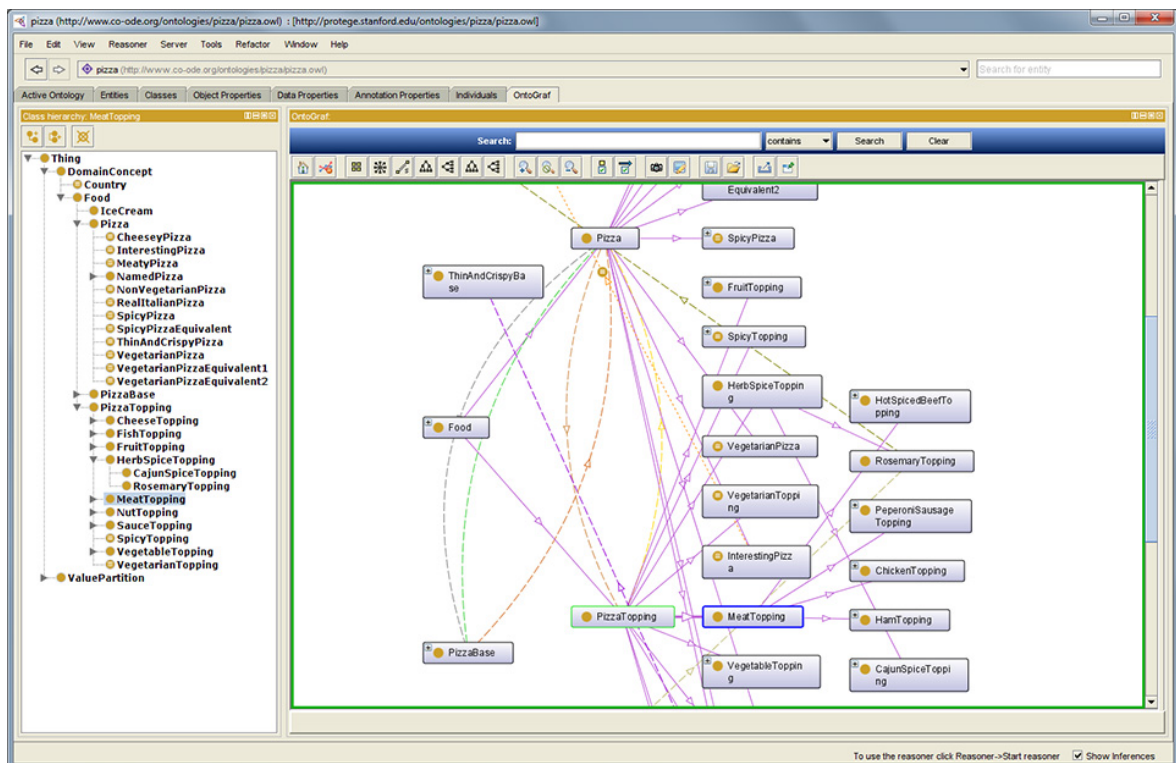


Figura 12: Captura de pantalla de Protégé 5.0.0.

La Figura 12⁷ muestra parte de la ontología descrita en Horridge (2011), un manual para la creación de ontologías con Protégé.

Opciones básicas de Protégé son la definición y descripción de clases y de propiedades tipo objeto (relaciones), siendo posible establecer dominio y rango de éstas, la definición y asignación a clases de propiedades tipo dato (atributos), el establecimiento de restricciones, y la creación de instancias y procesos de inferencia (automática) a partir de los anteriores. Gracias a OntoGraph (la pestaña activa en la Figura 12), es fácil visualizar clases primitivas e inferidas, así como los tipos de relaciones entre éstas. Los razonadores implementados analizan y muestran las inconsistencias de la ontología, es decir, la

⁷ <http://protege.stanford.edu/assets/img/screenshots/desktopprotege-screenshot-5.jpg> (último acceso: 20/05/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

verifican. Algunos aspectos se detallan en las siguientes figuras con la ontología Event.

Para la definición y descripción de clases y subclases se utiliza la pestaña *Classes*, que por defecto, se divide en tres partes (ver Figura 13):

La columna de la izquierda, *Class hierarchy*, permite crear o suprimir clases, así como establecer relaciones jerárquicas de clase. También es posible visualizar, además de las clases primitivas, las clases inferidas (*Class hierarchy (inferred)*). Para ver esas inferencias es necesario activar alguno de los razonadores implementados en el software.

En la parte superior derecha, *Annotations*, se pueden realizar anotaciones de varios tipos relativas a dichas clases. En la Figura 13 aparecen anotaciones *rdfs:label*, *comment*, *isDefinedBy* y *term_status*, y el uso de cada clase (*Usage*) dentro de la ontología.

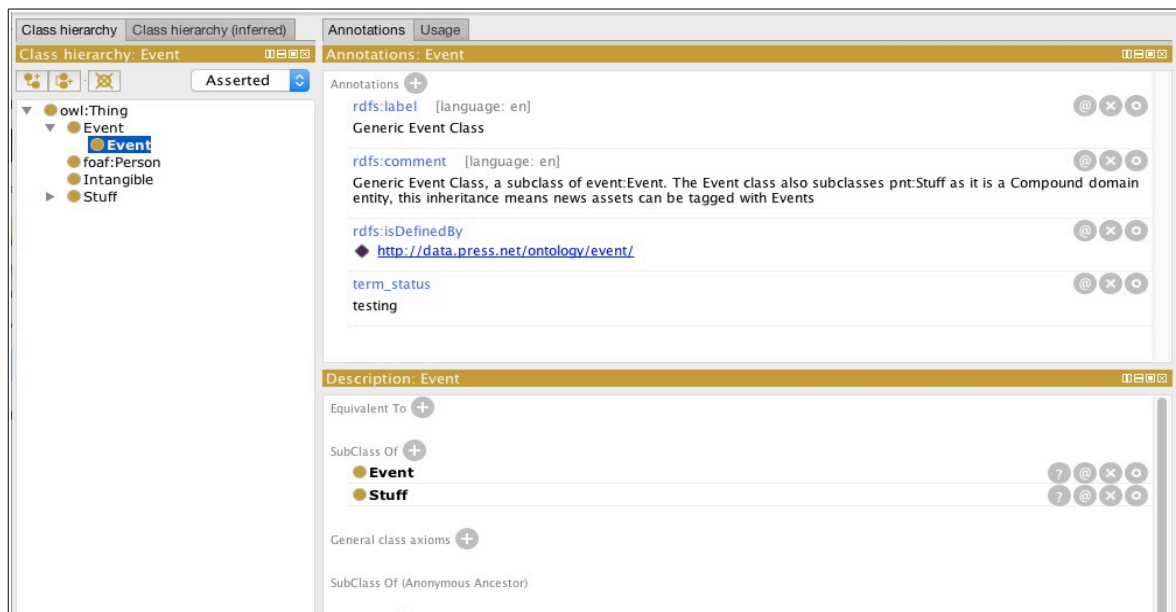


Figura 13: Pestaña para la definición de clases en Protégé 5.0.0.

Finalmente, en la parte inferior, derecha *Description*, se pueden establecer equivalencias entre clases (*Equivalent to*), definir una clase a partir de restricciones (Horridge, 2011, p. 37) (*SubClassOf*), ver las jerarquías de subclases (también en *SubClassOf*), etc.

Para la definición y descripción de relaciones existe una segunda pestaña, *Object Properties*, similar a la anterior a excepción de la parte derecha inferior (Figura 14): en la columna de la izquierda, se seleccionan las características de las propiedades, descritas en 2.2.1. COMPONENTES DE LA ONTOLOGÍA. Es aquí, en el panel *Description* donde se definen dominio (*domains*) y rango (*ranges*) de estas propiedades.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

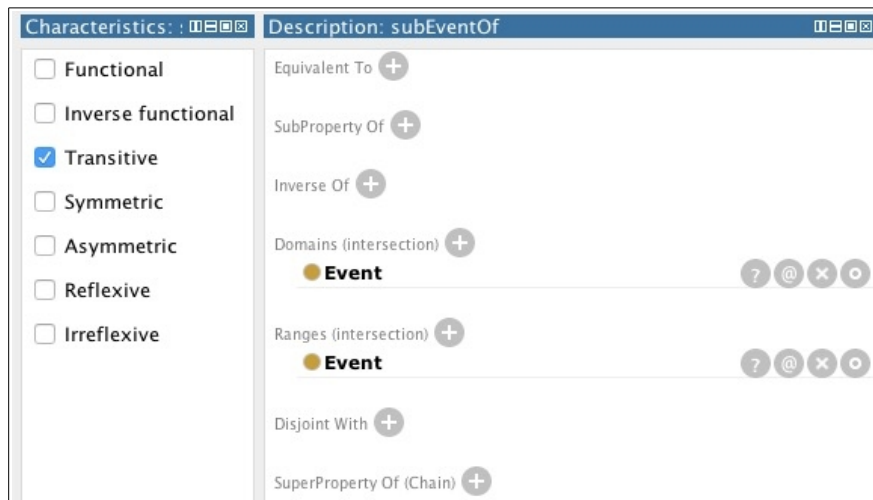


Figura 14: Pestaña para la definición de propiedades en Protégé 5.0.0.

El caso de las propiedades tipo datos (o atributos) es similar a las propiedades objeto, aunque la columna de características de la propiedad (*Characteristics*) sólo permite seleccionar una característica: *Functional*.

Finalmente, existe una última pestaña básica (a efectos de descripción del software en este documento), individuos (*individuals*), donde se definen las instancias de la ontología y se declaran las relaciones enumerativas entre clases e instancias, junto a otras particularidades.

En 2016 se desarrolló webProtégé⁸, un entorno web similar a la versión de Escritorio, aunque con opciones colaborativas: compartición y permisos, notas y discusiones, vigilancia y notificaciones. Entre las características de webProtégé, destacan⁹: soporte para OWL2, interfaz para una edición sencilla y personalizable, seguimiento de modificaciones e historial de revisiones y múltiples formatos de subida y descarga, entre los que se encuentran RDF/XML, Turtle y OWL/XML. Una de sus limitaciones es que carece de razonador.

Uno de los principios de la web semántica es la reutilización. En este sentido, Protégé permite importar elementos de otras ontologías. El siguiente apartado trata precisamente de los repositorios de ontologías, fuentes de las que extraer ontologías con vistas a su reutilización, destacando especialmente *Linked Open Data* (LOD).

8 <http://webprotege.stanford.edu> (último acceso: 06/09/2015)

9 <http://protegewiki.stanford.edu/wiki/WebProtege> (último acceso: 21/05/2017)

2.2.5. REPOSITARIOS DE VOCABULARIOS: *LINKED OPEN VOCABULARIES*

Existe un gran número de ontologías disponibles y continuamente se están desarrollando otras (Stuart, 2016, pp. 53-76). Sin embargo, uno de los problemas de la producción de ontologías en los últimos años es la ausencia de mecanismos y estándares para identificar y acceder a éstas (Hartmann, Palma y Gómez-Pérez, 2009). «Muchos investigadores piensan que tener bibliotecas de componentes reutilizables de ontologías debe mejorar el proceso de desarrollo de sistemas basados en conocimiento» (Martínez-Béjar et al., 1997). Así, surgen los repositorios o bibliotecas de vocabularios, con el propósito de facilitar la búsqueda, exploración, consulta y evaluación de ontologías (Dutta, Nandini y Shahi, 2015), entre otras herramientas.

Los registros de vocabularios, en general, están avanzando hacia un nuevo nivel, con desarrollos y proyectos (Méndez Rodríguez y Greenberg, 2012; Hartmann et al., 2009). Pero no todas las ontologías desarrolladas y disponibles han sido recolectadas por estos. Tampoco las diversas publicaciones identifican todos los repositorios existentes. Ello implica que, en el momento en que se plantea la reutilización de ontologías es necesario efectuar búsquedas que faciliten su acceso. Así, es habitual encontrar ontologías modeladas en una institución científica, que difunde sus investigaciones mediante artículos y no siempre en repositorios. De esta forma, la literatura científica puede resultar útil para acceder a ontologías que, estando disponibles en la red, no son localizables en una librería de ontologías o, incluso, mediante buscadores como Google o Swoogle.

Entre los diversos repositorios puede destacarse *Linked Open Vocabularies* (LOV)¹⁰, de la *Open Knowledge Foundation* (OKFN), que permite buscar vocabularios antes de definir otros para la descripción y representación de recursos (Pastor-Sánchez, 2017).

El registro LOV, creado por Bernard Vatant y Pierre-Yves Vandenbussch, tiene como finalidad proporcionar un acceso fácil a vocabularios (Méndez Rodríguez y Greenberg, 2012). La calidad de los metadatos y la diversidad de formas de interactuar con las combinaciones de datos hacen de LOV una de las bibliotecas de ontologías más amigables, aunque no la más grande (Stuart, 2016, p. 84).

Entre los requerimientos de inclusión de un vocabulario en LOV, se incluye la estabilidad de su URI y su disponibilidad en la web, la utilización de formatos estándar y buenas prácticas, metadatos y documentación de calidad, publicaciones identificables y confiables y una política de versiones adecuada¹¹. Antes de la inserción de un vocabulario, el equipo

10 <http://lov.okfn.org/dataset/lov/> (último acceso: 18/05/2017)

11 <http://lov.okfn.org/dataset/lov/about> (último acceso: 18/05/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Para cada vocabulario, LOV indexa a texto completo, clases, propiedades, tipos de datos e instancias. Además, recoge su URI, espacio de nombre, página web oficial, una breve descripción, idioma/s de los términos del vocabulario, autores y publicadores, lenguaje formal utilizado para la implementación del vocabulario y estadísticas sobre sus elementos. También se indica qué otros vocabularios están reutilizando el vocabulario (*incoming links*) y cuáles están siendo utilizados por éste (*outgoing links*). Adicionalmente, muestra una línea del tiempo que muestra sus diferentes actualizaciones. Finalmente, los vocabularios son clasificados por una o varias *tags*, por ejemplo: personas, metadatos, catálogos, geografía, eventos, gobierno, etc., lo que permite acceder a los vocabularios en función de su dominio o propósito. Algunos de estos datos pueden observarse en la Figura 16, que utiliza, a modo de ejemplo, la descripción de FOAF en LOV.

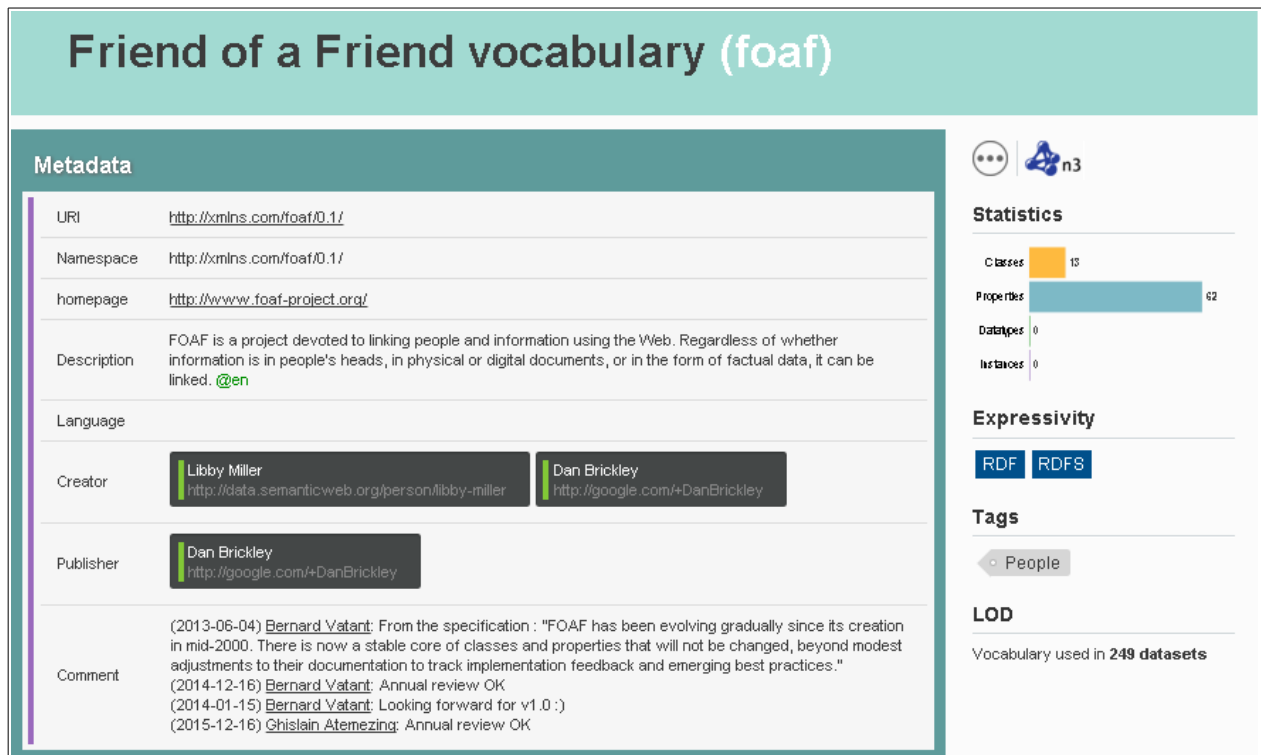


Figura 16: Datos recogidos por LOV para el vocabulario FOAF

Una vez seleccionado el vocabulario de interés, existen distintas posibilidades de reutilización, como el mapeo o la fusión de ontologías. LOV es utilizado en 4.4.3. MAPEO DE VOCABULARIOS como fuente de obtención de vocabularios de los que se plantea su reutilización en los modelos ONA y ONAPE.

2.2.6. REUTILIZACIÓN DE VOCABULARIOS: MAPEO

Es conveniente reutilizar tantos vocabularios como se pueda (Stuart, 2016, p. 75) ya que facilita la interoperabilidad entre sistemas. Esto se realiza mediante el mapeo entre vocabularios.

Se habla de mapeo (*mapping*) como una forma de vincular conceptos de dos ontologías, el proceso por el que se toman dos ontologías (una ontología fuente y una ontología objetivo) y se produce como salida un conjunto de relaciones semánticas entre las entidades de esas ontologías (Shvaiko y Euzenat, 2005). Este conjunto de elementos mapeados entre dos vocabularios es lo que los autores denominan alineamiento. Para Scharffe, Zamazal y Fensel (2014), sin embargo, el alineamiento de ontologías es la parte de la mediación de ontologías (*ontology mediation*) centrada en la construcción de una especificación de las correspondencias entre las partes que se superponen en dos ontologías. De forma similar, Bruijin et al. (2006) divide la mediación de ontologías en tres áreas:

- 1. Mapeo de ontologías:** centrado en la representación de las correspondencias entre ontologías, es una especificación de la superposición (*overlapping*) entre dos ontologías, empleando un lenguaje de mapeo. En este sentido, Martín Muñoz (2008) señala que, para indicar cómo dos vocabularios están relacionados «es preciso establecer algún mecanismo, en el lenguaje en el que la ontología esta expresada, que indique el modo en el que dos o más términos están relacionados». Este mecanismo es el **metadato**, información sobre un recurso existente.
- 2. Alineamiento de ontologías:** pone el foco en el descubrimiento automático y semiautomático de correspondencias entre dos ontologías fuente, empleando algoritmos basados en operadores de búsqueda de relación (*matching operator*). Su producto es la especificación de similitudes entre éstas.
- 3. Fusión de ontologías:** se interesa por la creación de una nueva ontología basada en x ontologías fuente. En este caso, no se definen elementos propios, todos son importados de vocabularios previos, entre los que existen algún tipo de relación.

Algunos trabajos consideran alineamiento y mapeo como sinónimos, por ejemplo Faria et al. (2013). También existen otros términos que reflejan ideas similares, como «integración, mezcla o traducción de ontologías» (Acurero, Bracho y Rincón, 2008). En esta memoria se emplea los términos mapeo y alineamiento como sinónimos. En cualquier caso, el objetivo de lo que, de forma más general Martín Muñoz (2008) denomina coordinación de ontologías, consiste en aprovechar el conocimiento expresado en éstas, estableciendo relaciones entre pares de ontologías, definiendo generalizaciones, especializaciones, conexiones, etc. y permitiendo la reutilización de información de una en otra (Lamarca Lapuente, 2013b; Martín Muñoz, 2008).

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

Scharffe et al. (2014) desarrollan una biblioteca de patrones comunes para ayudar a modelar correspondencias en un mapeo. Cada patrón representa una solución para un problema de alineamiento y se define a nivel de clase, relación o atributo. Un patrón especifica los tipos de relaciones entre elementos de dos ontologías como: equivalencia exacta (*equivalence*), incompatibilidad (*mismatch*), solapamiento o equivalencia parcial (*overlapping*), elemento más general (*more general*), y elemento más específico (*less general*) (Shvaiko y Euzenat, 2005). La ISO 25964-2, de interoperabilidad de los tesauros con otros sistemas de organización del conocimiento contempla estas relaciones entre vocabularios estructurados, aunque con diferentes denominaciones y clasificaciones. Así, se identifican 4 tipos de mapeo: equivalencia simple, equivalencia compuesta, jerárquico y asociativo, con distintos subtipos (García-Marco, 2017). La Tabla 7 recoge estos tipos de mapeo y subtipos. Por ejemplo, la clase de una ontología es asimilada con la intersección de dos clases de la segunda (*equivalent class intersection*, *subclass intersection*, *superclass intersection*). En granate, se muestra la denominación empleada por la ISO 25964-2, según García-Marco para el mapeo entre conceptos.

Tipo de mapeo		Patrones de correspondencia		
		A nivel de clase		A nivel de relación
Equivalence (<i>equivalencia exacta</i>) Equivalencia simple		Equivalent Class		Equivalent Relation
Overlapping (<i>equivalencia parcial</i>)		Class by Attribute	Class by Attribute Type Class by Attribute Occurrence Class by Attribute Value	Relation Negation
		Instance of Class		
		Class Intersection Equivalencia intersectiva	Equivalent Class Intersection Subclass Intersection Superclass Intersection	Inverse Relation
		Class to Attribute		Relation Transitive Closure
		Class to Relation		
		Class to Instance		
		Class Union Equivalencia acumulativa	Equivalent Class Union Subclass Union Superclass Union	Domain Restriction
Range Restriction				
Jerárquico	More general (<i>más general</i>)	Super-Class		Super-Relation
	Less general (<i>menos general</i>)	Sub-Class		Sub-Relation

Tabla 7: Tipos de relaciones entre elementos a mapear y patrones de alineamiento

Un ejemplo gráfico de patrón se muestra en la Figura 17, extraída del portal colaborativo de patrones de diseño de ontologías *Ontology Design Patterns (ODP)*¹³.

13 ODP: <http://ontologydesignpatterns.org/wiki/Category:AlignmentOP> (último acceso: 18/04/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

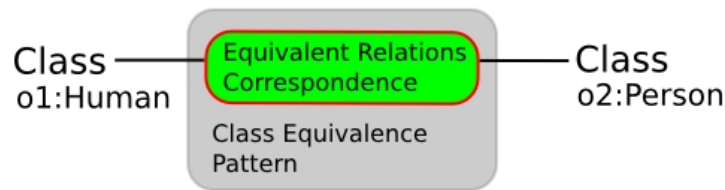


Figura 17: Patrón de equivalencia exacta entre clases de dos ontologías

Aplicando estos patrones a vocabularios concretos, Scharffe et al. (2014) muestran el alineamiento para la propiedad *name* de vCard y FOAF. De acuerdo con ésta, en la Figura 18, un nombre en vCard es modelado como una relación *vc:name*, que apunta a la clase *vc:N* que, a su vez, tiene tres propiedades: *vc:family-name*, *vc:given-name* y *vc:additional-name*, esto es, tres partes del nombre de una persona.

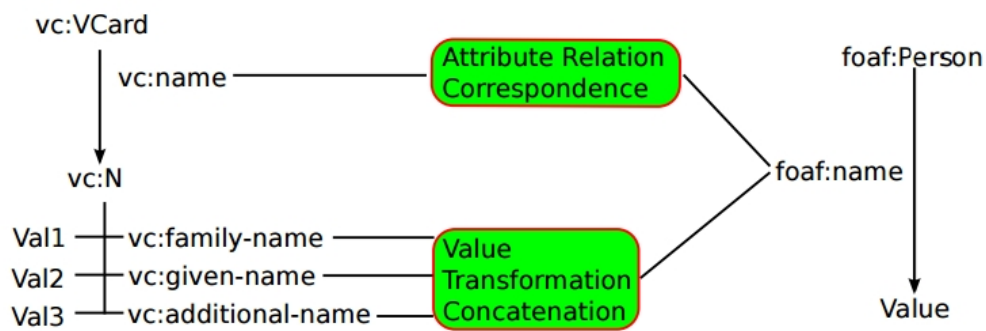


Figura 18: Alineamiento de la propiedad *name* de vCard y FOAF

En el alineamiento entre estos dos vocabularios, los autores procuran capturar la equivalencia entre la propiedad *foaf:name* y la relación *vc:name* y la concatenación de las tres propiedades señaladas.

Finalmente, hay que resaltar el desarrollo de técnicas de alineamiento automático, mediante algoritmos y software (véanse, por ejemplo Anam, Kim, Kang y Liu (2016) y Bruijin et al. (2006). De hecho, existe la iniciativa *Ontology Alignment Evaluation Initiative* (OAEI)¹⁴ para la evaluación de las fortalezas y debilidades de los sistemas de alineamiento, la comparación del rendimiento de las distintas técnicas, la facilitación de la comunicación entre desarrolladores de algoritmos y la mejora las técnicas de evaluación, contribuyendo así a mejorar la tarea de alineamiento de ontologías.

Existe un amplio rango de paquetes de software de alineamiento de ontologías, aunque muchos requieren de cierto nivel de experiencia, por su complejidad, como, por ejemplo, OnAGUI (Stuart, 2016, p. 122).

El siguiente apartado se centra en la indización como técnica para la descripción de documentos.

¹⁴ OAEI: <http://oaei.ontologymatching.org/> (último acceso: 18/04/2017)

2.3. LA INDIZACIÓN

El concepto indización se relaciona con el análisis de documentos, esto es, con la representación y descripción del contenido de los mismos, mediante conceptos contenidos en ellos o vocabularios controlados, con el propósito de guiar al usuario en la recuperación de documentos (Turpo Chaparro y Medina, 2013). La indización tiene por objetivo determinar el asunto de un documento y expresarlo mediante términos de indización (Nicolino y Ferneda, 2014; Salton y McGill, 1983, p. 53), que describen su contenido para hacer posible la recuperación de los documentos de un sistema. También se aplica en las ecuaciones de búsqueda, esto es, una pregunta planteada por un usuario ante una necesidad informativa. Para ello, «los términos de indización deben ser indicadores útiles del asunto de los textos a los que se aplican» (Raber, 2003, p. 147).

Un concepto relacionado es indexación, un «proceso por el cual el buscador va rastreando [un sitio] y al mismo tiempo va incorporando a su base de datos el contenido de esas URLs (...) para elaborar un índice» (Turpo Chaparro y Medina, 2013). También se define como proceso por el que se crean índices en un documento mediante el marcado de palabras no vacías, para que puedan ser utilizadas en la recuperación de información en un sistema (Gil Leiva, 2012, pp. 72-77). Hay quien considera indización e indexación como sinónimos: el Diccionario panhispánico de dudas da prioridad a «indexar» ya que el término español «indizar», derivado del sustantivo español índice, es de uso mucho menos frecuente (Rodríguez Yunta, 2006).

Cavalcanti (1978, p. 3) sitúa el origen de la indización como instrumento para el almacenamiento y recuperación de información siglos atrás, en la Biblioteca de Alejandría y en la clasificación de Calímaco y, posteriormente, en las bibliotecas de la Edad Media. La segunda Guerra Mundial (1939-1945) y la posguerra, con la llamada *explosión de la información*, evidenciaron los problemas del aumento de la producción bibliográfica: su falta de control y la dificultad para recuperarla. Así es cómo se fueron desarrollando sistemas más adecuados y rápidos que facilitaron el acceso a los documentos.

Con frecuencia, la indización es descrita como un proceso de dos pasos: análisis del documento para determinar el asunto tratado y traducción del asunto a un conjunto de términos. Anderson y Pérez-Carballo (2001a), Chowdhury (2010, pp. 100-101) y Lancaster (2004, p. 9) amplían este proceso a cinco etapas: 1) Análisis de la materia: decidir cuál es el asunto del que trata el documento; 2) Identificación de las palabras clave: representar ese asunto mediante términos adecuados y que, además, se puedan ajustar a las ecuaciones de búsqueda de los usuarios de un sistema documental; 3) Estandarización de las palabras clave; 4) Elección de un sistema de indización: por precoordinación o por postcoordinación; 5) Presentación de los asientos. Como resultado, obtenemos la representación del contenido de un documento, contra la que se interroga un sistema

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

documental para acceder a un grupo de documentos relacionados. Por este motivo, Ingersoll, Morton y Farris (2013, p. 41) definen indización como «el proceso del motor de búsqueda de hacer uno o más documentos “buscables”».

Este resultado puede variar en función del tipo documental, pero también en función del tipo de aproximación con el que se indiza: 1) orientada al usuario (*user-oriented approach*), en la que se tiene en cuenta no sólo el documento que se analiza, si no cuáles pueden ser las necesidades informativas de los usuarios; 2) orientada al documento (*document oriented approach*), en la que sólo se considera el documento y lo indicado por los autores. Una variante de la primera aproximación pone el foco de atención en el dominio (*domain-centered approach*) que, además del usuario, presta atención a los objetivos, propósitos y actividades del dominio del documento. Una variante de la segunda aproximación se centra en el documento (*document-centered approach*), pero también tiene presente las preguntas que el/la indizador/a cree que se pueden plantear para la recuperación del documento. En esta tesis, la fase de indización humana tiene en cuenta esta última variante.

Finalmente, la efectividad de una indización, en términos de recuperación de documentos, se ve influida por cuestiones como la exhaustividad, la especificidad y la corrección de la indización, así como por la exhaustividad y precisión de la recuperación en sí, como se detalla posteriormente.

En el siguiente apartado se describen indización humana y automática de documentos, ambas utilizados en esta tesis.

2.3.1. INDIZACIÓN HUMANA E INDIZACIÓN AUTOMÁTICA

La indización humana (también llamada manual o intelectual) es el proceso cognitivo (Anderson y Pérez-Carballo, 2001a) por el que una persona determina qué conceptos representan el contenido de un documento o ecuación de búsqueda. La indización automática, también llamada indización inteligente (Lamarca Lapuente, 2013a), hace uso de técnicas estadísticas y/o lingüísticas para la asignación de dichos conceptos.

El consenso general es que los indizadores humanos perciben (leen, ven, examinan, escuchan) un texto, lo interpretan y lo describen de acuerdo con su versión del mensaje, teniendo en cuenta una serie de reglas y patrones (Anderson y Pérez-Carballo, 2001b). El bagaje de quien indiza (formación, conocimiento previo, etc.), documento y contexto también influyen en la indización.

Otro factor que influye es la subjetividad. Los vocabularios controlados son inevitablemente subjetivos, reflejan la visión del mundo de sus creadores en un momento dado (Stuart, 2016, p. 7). La subjetividad afecta no sólo al vocabulario, también a los mismos documentos: «diferentes culturas, diferentes lenguajes y diferentes

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

interpretaciones del mismo texto pueden hacer que el mejor ingeniero se pregunte qué implementar» (Ingersoll et al., 2013, p. 10). La Tabla 8, a partir de Gil Leiva (2012), Gil Leiva (2008, pp. 78-96), Lancaster (2004), Narukawa, Gil Leiva y Fujita (2009), recoge algunos de estos factores de influencia. Es conveniente tener en cuenta que la indización humana, como cualquier proceso intelectual, puede estar influida por la cultura y experiencias personales de quien indiza, y puede que no sea fácilmente trasladable a otras culturas. Ahora bien, la máquina también tiene una cultura, la impuesta por sus programadores (Anderson y Pérez-Carballo, 2001a).

Elemento	Indización humana	Indización automática/semiautomática
<i>Indizador/a</i>	Formación del/a indizador/a	No influyen factores externos, pues una máquina siempre indizará de la misma forma. Sólo variará cuando se alteren los parámetros de análisis del sistema. Esto no es aplicable al momento en que se crean el sistema de indización pues, como todo producto humano, puede verse influido por su productor
	Conocimiento previo	
	Capacidad de realizar inferencias a partir de la lectura, permitiendo relacionar el contenido con el contexto en que se enmarca, crear un significado global, etc.	
	Grado de profesionalismo	
	Motivación del/a indizador/a	
<i>Documento a indizar</i>	Características del objeto indizado	Características del objeto indizado
<i>Contexto de la indización</i>	Condiciones en que se realiza la indización: herramientas, política del centro, etc., objetivos de la indización, usuarios para quienes está destinada	Condiciones en que se realiza la indización: herramientas, política del centro, etc., objetivos de la indización, usuarios, etc., variará cuando se alteren los parámetros de análisis del sistema

Tabla 8: Cualidades de la indización humana y automática

Una de las líneas más investigadas en este tipo de descripción de documentos es el Procesamiento de Lenguaje Natural (PLN). Éste puede definirse como «un intento de simular el comportamiento lingüístico humano, de manera que el sistema de signos que constituye la lengua sea adquirido y procesado por el computador, siendo éste capaz de reconocer, comprender, interpretar y generar lenguaje humano» (Ruíz-Martínez, 2011, p. 41). El PLN forma parte de la evolución de las técnicas de indización automática (Lamarca Lapuente, 2013a), y pone el foco de atención en el tratamiento del lenguaje natural. Se pueden considerar tres etapas de desarrollo:

1. Años sesenta. Las palabras se entienden como objetos. Se introducen métodos estadísticos y probabilísticos. En esta etapa ya se utilizan listas de palabras vacías¹⁵ y se analiza la frecuencia de aparición de cada palabra en los textos, eliminando las muy frecuentes y las escasas.

¹⁵ Palabras vacías son (pequeñas) palabras que no son tenidas en cuenta en el proceso de indización porque se supone que no tienen significado (Keyser, 2012, p. 67)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

2. Años setenta. Se utiliza el análisis lingüístico como medio que reduce la ambigüedad de los términos en la oración.
3. Actualidad. Se lleva a cabo una indización (automática) inteligente, que combina estadística, lingüística y bases de conocimiento para la extracción de los conceptos presentes en los documentos. Por base de conocimiento se entiende «un tesoro enriquecido con información morfológica, sintáctica y semántica, esto es, una ontología» (Lamarca Lapuente, 2013a). Obviamente, puede aplicarse a otro tipo de vocabularios.

Las fases del Procesamiento de Lenguaje Natural (PLN), entendido como un tipo de indización automática, son según Ruíz-Martínez (2011, pp. 43-44) y Salton y McGill (1983, pp. 259-260) las que se resumen a continuación:

1. **Preprocesamiento textual:** segmenta el texto en *tokens* o palabras y otras unidades (oraciones y párrafos) para su análisis posterior en base al reconocimiento de espacios y signos ortográficos, como puntos y comas.
2. **Procesamiento léxico-morfológico:** descompone las palabras en función de sus componentes para determinar su categoría gramatical. Consiste en la detección de las unidades que componen la palabra: lexema y morfemas (prefijos, interfijos, sufijos) y la identificación de las categorías gramaticales de las palabras. A este nivel se suele aplicar un *stemming* o reducción de un término a su raíz, para lo que se emplean listas de sufijos (Ingersoll et al., 2013, p. 25; Salton y McGill, 1983, p. 71), o de *lematización* o reducción del término a su forma canónica.
3. **Procesamiento sintáctico:** identifica las funciones de las palabras en la oración. El resultado es una «estructura correspondiente a las categorías sintácticas formadas por cada una de las unidades léxicas que aparecen en la oración» (Sosa, 1997). Para este fin, se utiliza una gramática que describe las estructuras de las oraciones.
4. **Procesamiento semántico:** establece relaciones semánticas entre las estructuras textuales para determinar su sentido. Para ello, se aplican técnicas de reconocimiento y clasificación de entidades nombradas y su fin es «resolver de manera automática algunos tipos de sinonimia, polisemia, anáfora o elipsis», frecuentes en el lenguaje natural (Gil Leiva, 2008, p. 349).
5. **Procesamiento pragmático:** establece relaciones entre un enunciado y el contexto comunicativo en el que se produce. Es un proceso muy complejo e implica la utilización de algoritmos que «recogen las relaciones contextuales para poder comprender el contexto del discurso» (Lamarca Lapuente, 2013a).

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

El PLN puede utilizarse tanto para extraer términos de textos en lenguaje natural, para incorporar elementos a una ontología o tesoro, como para anotar o marcar semánticamente (indizar) textos mediante términos de una ontología o tesoro (Stuart, 2016, p. 23). La indización humana vendría a coincidir con el PLN en las dos últimas etapas de procesamiento, esto es, procesamiento semántico y procesamiento pragmático.

Indización humana y automática cuentan con argumentos a favor y en contra. A la indización humana se le atribuye un tiempo de procesamiento y costes elevados, falta de detalle, obsolescencia, errores, subjetividad en la asignación de los términos y problemas de consistencia. La indización automática trata de imitar la manual pero sin estos problemas (Abadal y Codina, 2009, p. 65). A ésta se le achaca la falta de una visión coherente de los términos de indización, la no solución de los problemas de sinonimia y sus variantes, tampoco de ortografía, el hecho de que no tiene en cuenta el contexto del documento ni permite la búsqueda de materias relacionadas (Keyser, 2012, pp. 40-56). Ambas formas de procesamiento hacen frente, por tanto, a problemas del lenguaje natural, entre otros, destacando especialmente la ambigüedad, que surge «cuando no podemos determinar unívocamente cuál es el significado de algo» (Escolano Ruiz, Cazorla Quevedo, Alfonso Galipienso, Colomina Pardo y Lozano Ortega, 2003, p. 205). Dependiendo del contenido y del propósito del PLN puede ser categorizada más como proceso semiautomático que como automático (Stuart, 2016, p. 23).

Sin embargo, es indudable que el futuro pasa por el desarrollo de técnicas de indización automática, pues el crecimiento de la información disponible en Internet hace muy difícil el mantenimiento de procesos netamente intelectuales (Lamarca Lapuente, 2013a). De hecho, en el ámbito periodístico, el crecimiento de la información es tal que la reducción de los tiempos de producción y difusión de noticias ha propiciado una mayor aceptación de «métodos automáticos de análisis y visualización de datos que apoyen la creación, en un tiempo crítico, de conocimiento» (Tassilo Pellegrini, 2012). Así, parte del PLN se está orientando hacia la descripción y recuperación de información, sin suplir el razonamiento humano, aunque sí lo asiste en buena medida. Mientras estos sistemas son perfeccionados, el recurso a una combinación de técnicas más o menos automáticas y a la revisión humana puede ser una solución adecuada.

La Figura 19 muestra dos procesos, descripción de documentos y modelado de vocabularios. En el primero, los documentos primarios de un dominio son indizados empleando herramientas de representación del conocimiento, generando documentos secundarios que los representan. En el segundo, tanto los documentos primarios como los secundarios son fuentes de datos para el modelado y construcción de vocabularios como las ontologías, también empleando la indización como herramienta para el aprovisionamiento de elementos del vocabulario. La indización, en el centro de ambos

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

procesos, puede ser automática o humana. Las dos últimas etapas del PLN, procesamiento semántico y procesamiento pragmático, serían el equivalente al razonamiento humano durante la indización manual, aunque con contribuciones diferentes (Anderson y Pérez-Carballo, 2001b).

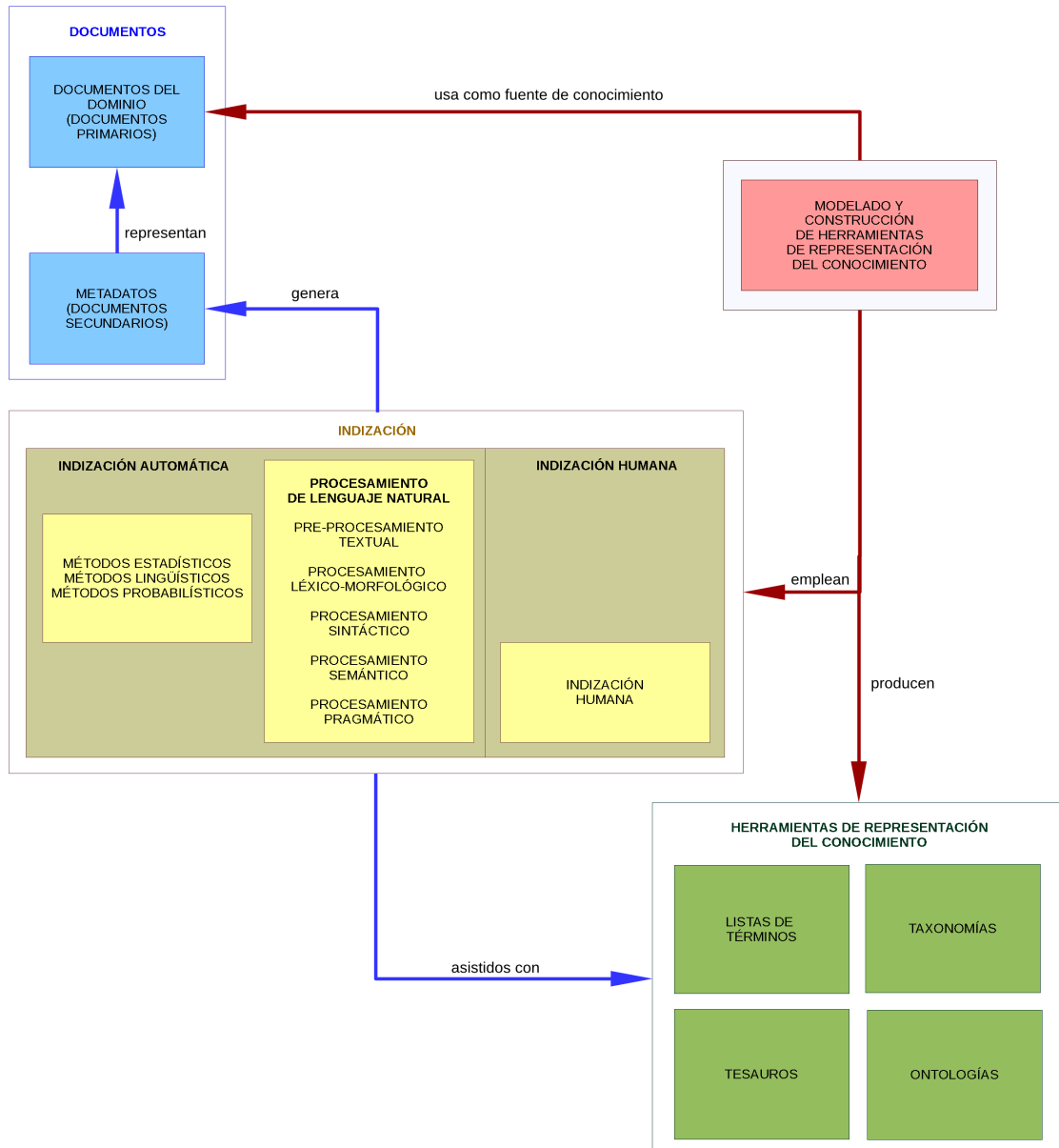


Figura 19: Relación entre descripción de documentos-indización-modelado de vocabularios

2.3.2. EXTRACCIÓN DE TÉRMINOS CON N-GRAMAS: e-TERMOS. ANÁLISIS MORFOLÓGICO: GRAMPAL

Una de las técnicas de indización automática es n-gramas, utilizada también para la recuperación de información. En general, en este modelo se puede pensar en el texto como en una ventana de contexto (*windows of context*) cuyo tamaño depende de cuánto contexto se tendrá en cuenta durante el análisis (Ingersoll et al., 2013, pp. 30-31). Esa ventana trabaja, bien con caracteres, bien con términos, a los que se denomina gramas. Se trata de analizar el texto como una secuencia de elementos, a la que le preceden y le siguen otras.

Definido el tamaño de la ventana, se calculan las semejanzas entre uno o más gramas. Para ello, se utiliza la distancia de edición (*edit distance*), una suma de operaciones de edición necesarias para transformar un carácter en otro. Esas operaciones son habitualmente inserción, supresión o sustitución. Por ejemplo, para transformar la cadena *tamming test* en *taming text* se requiere la supresión del carácter *m* y la sustitución de *s* por *x*, por lo que la distancia de edición es 2. Esto es conocido como distancia Levenshtein (*Levenshtein distance*) (Ingersoll et al., 2013, p. 89), aunque existen otras variantes.

Para este cálculo se pueden utilizar distintas fórmulas, como el Coeficiente de similitud de Dice (C_d), con un resultado comprendido entre 0 (ausencia de semejanza) y 1 (semejanza total): $C_d = (2 \times C) / (A + B)$, donde: A = N° de bi-gramas únicos del 1º término; B = N° de bi-gramas únicos del 2º término; C = N° de bi-gramas únicos que comparten ambos términos. En la Figura 20 se comparan los términos *arma* y *armamento*, que tienen un coeficiente de similitud de 0,75. Aquí, cada grama está compuesto por un carácter y el tamaño de análisis es el bigrama, compuesto por dos gramas.

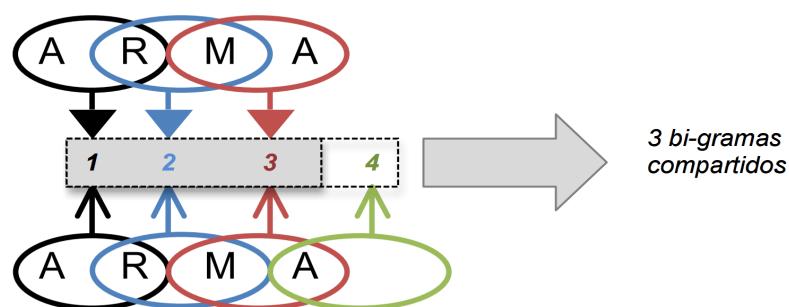


Figura 20: Comparación de términos con n-gramas

Una de las ventajas de este modelo, entre otras, es que puede aplicarse a cualquier conjunto de textos, independientemente del idioma (Martínez-Fernández, García-Serrano, Martínez y Villena, 2004). Sin embargo, esto sólo es posible tratando los *corpora* por lengua, de forma separada. Si además se utilizan listas de palabras vacías se reduce en buena medida el número de elementos a comparar (Anderson y Pérez-Carballo, 2001b).

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

2.3.2.1. EL SOFTWARE E-TERMOS

Entre las aplicaciones que usan n-gramas se encuentra e-Termos (*Termos Eletrônicos*)¹⁶, creado por el brasileño *Centro Nacional de Pesquisa Tecnológica em Informática para a Agricultura - Embrapa Informática Agropecuária* (CNPTIA) en colaboración con las Universidades de São Paulo (USP) -Brasil- y Federal de São Carlos (UFSCAR) -Brasil-.

Es un ambiente computacional colaborativo web de acceso libre dedicado a la gestión terminológica (Oliveira, 2009a). e-Termos funciona por etapas y, aunque no todas están desarrolladas, permite aplicar la técnica sobre un corpus de documentos, desde 1 grama (un término) a 7 (siete términos). De esta forma, la co-ocurrencia de 2 o más términos también es tenida en cuenta en esta investigación. Al final del proceso, si se aplicaran todas las etapas, se obtiene una herramienta de representación del conocimiento que aplicar sobre documentos primarios. Algunas características de e-Termos son:

1. *Interfaz amigable*. Conociendo el funcionamiento de algunas de las técnicas de PLN implementadas, se puede utilizar sin conocimientos matemáticos profundos.
2. *Trabajo por etapas*, facilitando la aplicación de técnicas PLN y desarrollo de herramientas de web semántica en distintos grados (ver Tabla 9). En esta tesis se utilizan las etapas 2 y 3.

ETAPA	DESCRIPCIÓN DEL FUNCIONAMIENTO
1 y 2	Configuración de los <i>corpora</i> sobre los que aplicar las técnicas de PLN
3	Aplicación de métodos de extracción automática de unidades o conjuntos léxicos que pueden constituir una unidad terminológica mediante técnicas n-gramas. El resultado es una o varias listas de candidatos a términos válidos
4	Creación de instancias a integrar en una ontología preexistente a partir de los candidatos a términos obtenidos en la etapa anterior
5	Creación de un banco de datos terminológico tomando como referencia los términos y la ontología de la etapa 4
6	Presentación del producto final

Tabla 9: Etapas de procesamiento de la información en e-Termos

3. Permite desechar *palabras vacías* añadiendo al sistema un fichero con un listado de dichas palabras (conocido en inglés como *negative dictionary* o *stop list*).
4. Trabaja en distintos idiomas, aunque de forma separada.
5. Ofrece distintos resultados en función del *tipo de documento*. Para ello, cuando se suben los *corpora* de trabajo (etapa 2) se define el género (científico, científico de divulgación, informativo, instruccional, publicitario, técnico-administrativo u oficial) y el tipo textual (artículo, ley, noticia, etc.).

16 <https://www.etermos.cnptia.embrapa.br/index.php> (último acceso: 21/06/2017)

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

6. Permite establecer una frecuencia mínima a partir de la cual desechar resultados. Si el sistema determina que una frecuencia inferior a 2 ocurrencias es irrelevante, mostrará aquellos términos que aparezcan al menos 3 veces en el conjunto de textos que se analizan. Este límite puede ser modificado manualmente.
7. En combinaciones de dos o más gramas muestra la frecuencia de aparición de cada grama y de su combinación.

Una descripción más detallada del software puede encontrarse en Oliveira (2009b).

2.3.2.2. ANÁLISIS MORFOSINTÁCTICO CON GRAMPAL

Posteriormente, es posible utilizar analizadores morfosintácticos que asistan el análisis de los términos detectados por e-Termos. Una herramienta sencilla en castellano y disponible en línea es Grampal¹⁷. Este etiquetador morfosintáctico fue desarrollado por Moreno Sandoval y Guirao Miras en el seno del Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid. La Figura 21 muestra un ejemplo de uso del software.

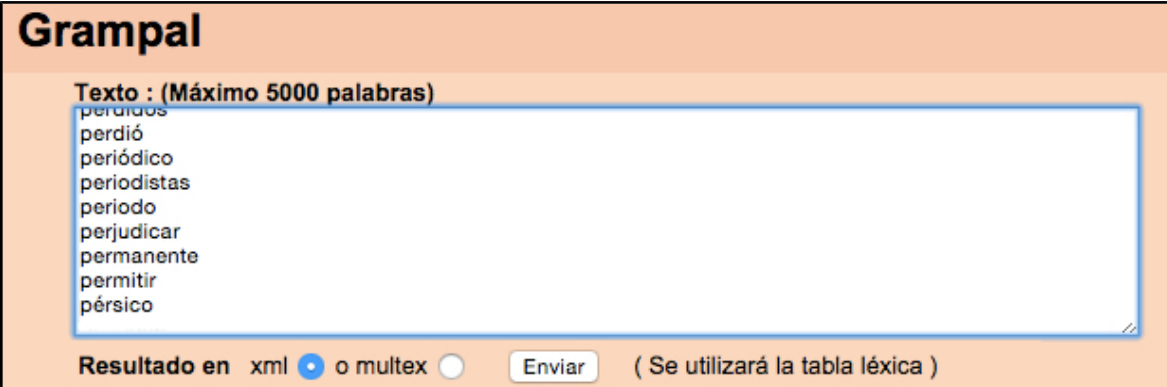


Figura 21: Ejemplo de uso de Grampal

Al introducir un texto completo en el sistema, se obtiene una lista de resultados compuesta por tres partes: palabra original (en el texto), forma canónica y morfología. Es decir, aplica un proceso de lematización y no de *stemming*. Así, se pueden agrupar aquellos términos que tienen un mismo origen y reducir el número de elementos a revisar. Además, el análisis morfosintáctico puede facilitar la definición de clases, relaciones y atributos en una ontología, de acuerdo con la función del término en la oración. En general, la lematización es mejor que el *stemming*, pues «en una sola unidad lingüística pueden aglutinarse [varias categorías gramaticales. Además, existen múltiples casos de lexemas irregulares]». Por otro lado, para el etiquetado gramatical «la reducción a la raíz tampoco es suficiente, ya que un conjunto de palabras que compartan el mismo lexema no tienen por qué pertenecer a la misma categoría gramatical» (Ruíz-Martínez, 2011, pp. 46-47).

¹⁷ <http://cartago.llf.uam.es/grampal/grampal.cgi> (último acceso: 21/05/2017)

2.3.3. INDIZACIÓN DE NOTICIAS. LA GRAMÁTICA DE CASOS

Una noticia es un documento que relata objetivamente un hecho susceptible de ser de interés público, por ser novedoso, destacable y/o actual. Desde el punto de vista de su estructura, ésta se divide en, al menos, titular, subtítular y cuerpo del texto. El titular, una síntesis que realiza el productor de una noticia «con el fin de atraer la curiosidad del lector» (García Gutiérrez, 2014), permitiría identificar el tema, y otros elementos como la acción principal o quién la ejecuta (Debnath y Giles, 2005; Gil Leiva, 2012, p. 40). Sería, pues, una frase que resume el contenido de la noticia (Kodama, Ozono, Shintani y Aosaki, 2008). Cada parte de la noticia responde a varias de las cuestiones del Paradigma de Lasswell¹⁸ : qué (*what*), quién (*who*), cuándo (*when*), dónde (*where*), cómo (*how*) y por qué (*why*). Estos elementos se muestran en la Figura 22, basada en (Gil Leiva, 2012, p. 40).

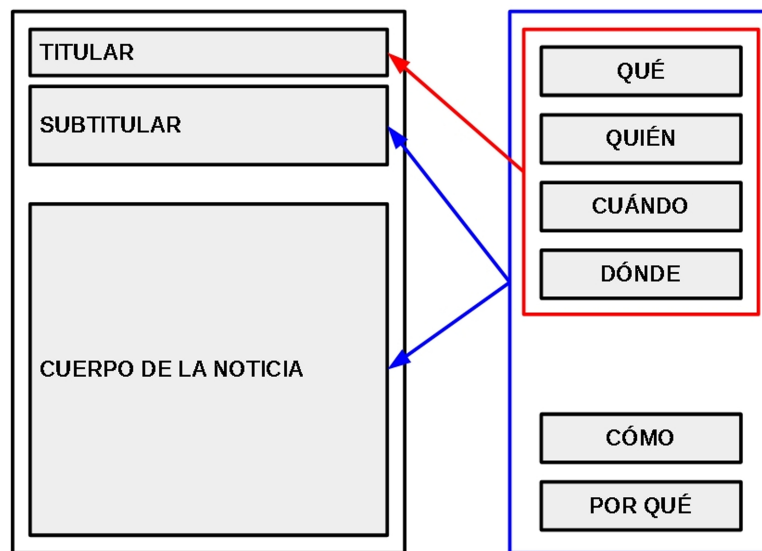


Figura 22: Paradigma de Lasswell y estructura de la noticia

Identificar y extraer las respuestas a los elementos del Paradigma de Lasswell permitiría indizar, en mayor o menor grado, una noticia (Gil Leiva, 2012, p. 40), esto es, representar su contenido mediante términos de indización.

Otra aproximación, de origen lingüístico, es la Gramática de Casos de Fillmore (1967) (*Case Grammar*). Consiste en el análisis de una oración en función de las relaciones sintáctico-semánticas entre los “casos” universales que la componen. Los casos son «seres, cosas, hechos y situaciones extralingüísticas comunes a la experiencia humana» (García-Hernández, 1986) que facilitan la representación del contenido de un documento cualquiera. Casos típicos para los sustantivos son agente, objeto, instrumento u objeto indirecto (Salton y McGill, 1983, p. 263). En esta línea, García Gutiérrez (2014) realiza un «inventario exhaustivo» de las preguntas necesarias para obtener los datos factuales de una

18 Inicialmente, estas cuestiones se centraban en los actos de comunicación de masas.

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

noticia y propone que sean respondidas a través de la Gramática de Casos. Este método no es nuevo, normas como los estándares BS 6529:1984 e ISO 5963:1985, ya planteaban conjuntos de preguntas a “responder” por los textos (Anderson y Pérez-Carballo, 2001a).

Para la definición de las 12Ws, García-Gutiérrez, basándose principalmente en Fillmore (1967) construye macroproposiciones, a las que aplicar dicha formulación (ver Tabla 10). Una macroproposición es la “reescritura” (y, en su caso, sustitución) y enriquecimiento del «título en un mismo enunciado canónico del tipo: sujeto-acción-predicado con los datos que el propio texto original explícitamente proporcione».

Ws planteadas por García Gutiérrez	Respuestas a Ws
¿Quién/Qué?	Sujeto que precede a la acción
¿A/Contra quién?	Receptor/es directo/s de la acción
¿Qué cosa es objeto de la acción?	Objeto beneficiado o perjudicado por la acción
¿Mediante qué se realiza la acción?	Procesos no tangibles especializados u ordinarios
¿A través de qué se realiza la acción?	Instrumental. Objeto tangible a través del cual se ejecuta la acción, precisan la mediación o manipulación por sujetos
¿Por qué se realiza la acción?	Causal. Extensible a motivos y antecedentes
¿Para qué se realiza la acción?	Intenciones y objetivos no alcanzados
¿Con qué consecuencias se realiza la acción?	Resultados alcanzados y constatados
¿Cuándo se realiza la acción?	Fecha, tiempo, periodo
¿Dónde se realiza la acción?	Lugar. Espacios genéricos o específicos
¿Con quién/qué se realiza la acción?	Asociativo. Sujeto colaborador, no protagonista, acompañante
¿En qué estado se realiza la acción?	Aparece junto al elemento que matiza

Tabla 10: 12W de una noticia, según García Gutiérrez (2014)

En la Figura 23, basada en García Gutiérrez (2014), Gil Leiva (2012, pp. 41-42) y Fillmore (1967, p. 46) y la Tabla 10, se muestra la relación entre el Paradigma de Lasswell, los casos de Fillmore y las «funciones conceptuales asociadas a preguntas pronominales» de García Gutiérrez (2014) (en adelante, “casos de García Gutiérrez”). Este trabajo es un buen punto de partida pues, como el autor indica, su procedimiento «abre un vasto campo de investigación (...) en materia de lenguajes de representación, construcción de thesauri, ontologías y epistemografías en el campo de la información de actualidad». En esta investigación los casos de García Gutiérrez son los elementos del modelo ONA.

Además, estas 12W pueden concebirse como cuestiones de competencia informal, una forma de especificar los requisitos a los que debe responder la ontología, como plantean Noy y McGuinness (2001), Grüninger (1995) y Fox o Mendonça y Almeida (2014), ya descritos, por ejemplo.

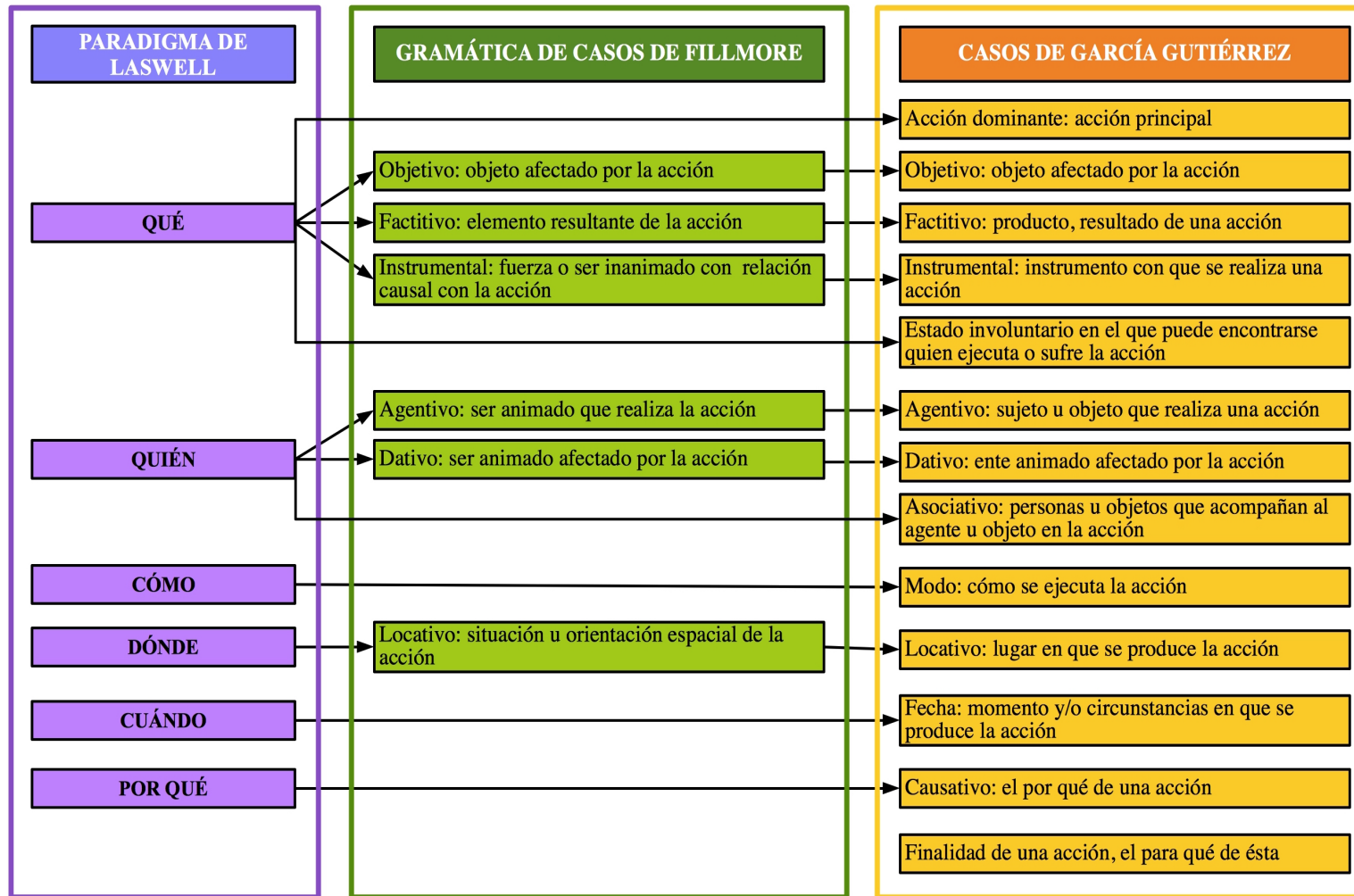


Figura 23: Relación Lasswell, Fillmore y García Gutiérrez. Basada en García Gutiérrez (2014)

2.3.4. EVALUACIÓN DE LA INDIZACIÓN

En la evaluación de los resultados de una indización, bien en la indización en sí o en la recuperación de información destacan distintas características. Éstas han sido analizadas por varios autores, entre los que pueden destacarse Cleveland y Cleveland (2013, pp. 158-160), Gil Leiva (2012, pp. 78-84), Chowdhury (2010, pp. 99-100), Martínez Méndez (2004), Martínez Méndez (2002), Salton y McGill (1983) y Cavalcanti (1978, pp. 56-57). Éstas permiten determinar en qué medida los términos asignados a un documento representan el contenido del mismo; también si éste responde adecuadamente a la pregunta formulada por un usuario. A continuación, se resaltan algunas características de la indización:

- **Relevancia** (*relevance*): es la medida de la efectividad de un contacto entre quien envía un mensaje y su receptor en un proceso de comunicación. Los criterios de relevancia, en general, tienen en cuenta el contenido informativo de un documento, su interpretación por un usuario y la motivación de éste cuando accede a ese documento. Relevancia es la relación de un documento con la necesidad de un usuario, que dicho documento ayuda a resolver (Moens, 2000, p. 13).
- **Exhaustividad en la indización y especificidad** (*specificity*) son dos aspectos relacionados. La primera implica reconocer y asignar al documento todos los conceptos que lo representan. El segundo, en ocasiones llamado precisión de la indización (Cleveland y Cleveland, 2013, p. 159), implica dar preferencia a los conceptos más específicos para representar el contenido de un documento en lugar de utilizar otros más generales.
- **Corrección** tiene que ver con evitar errores en la asignación de conceptos por dos vías, 1) inclusión de aquellos elementos que no representan el contenido de un documento y 2) omisión de conceptos que sí lo hacen.
- **Consistencia en la indización**, según Gil Leiva (2012, pp. 81-84) y Salton (1975, p. 99) mide el nivel de concordancia entre dos o más indizaciones, que pueden ser efectuadas por dos o más personas o máquinas (inter-consistencia) o por la misma persona o máquina en momentos diferentes (intra-consistencia). Para ello se calcula el coeficiente de similitud entre las dos indizaciones. Cuanto más se aproximen las indizaciones, mayor será la consistencia.
- **Exhaustividad en la recuperación** (*recall*) y **precisión** (*precision*) relacionan la cantidad de conceptos asignados a un documento y la ecuación planteada para su recuperación en un sistema. En una búsqueda, la exhaustividad tiene en cuenta cuántos documentos relevantes son recuperados en relación con el total de documentos relevantes del sistema (ver Figura 24). La precisión valora la cantidad

2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA

de documentos relevantes recuperados respecto del total de documentos recuperados (Salton y McGill, 1983, p. 55; 164) (ver Figura 25). Esto se puede resumir en: «un sistema ideal dará al usuario todos los documentos que necesite [exhaustividad], y no más [precisión]» (Cleveland y Cleveland, 2013, p. 159). Ambas medidas son inversas: a mayor exhaustividad, menor precisión, y viceversa.

$$\text{Exhaustividad} = \frac{\text{Total de documentos relevantes recuperados}}{\text{Total de documentos relevantes del sistema}}$$

Figura 24: Fórmula para el cálculo de la exhaustividad en una búsqueda

$$\text{Precisión} = \frac{\text{Total de documentos relevantes recuperados}}{\text{Total de documentos recuperados}}$$

Figura 25: Fórmula para el cálculo de la precisión de una búsqueda

La exhaustividad de una recuperación obtendrá mejores resultados mediante la utilización de términos más genéricos. Por el contrario, la precisión se verá recompensada con el empleo de términos específicos o con la combinación de varios términos (Salton, 1975, p. 92; 100). Es decir, cuanto mayor sea la exhaustividad de la indización, mayor será la exhaustividad de la recuperación (menor silencio informativo); cuanto mayor sea la especificidad de la indización, mejor será la precisión de la recuperación de información (menor ruido informativo) (Salton y McGill, 1983, p. 162).

Ambas medidas se suelen relacionar mediante diversas fórmulas, por ejemplo, la medida o valor F (*F measure* o *F score*) (Figura 26).

$$F = 2 \times \frac{\text{Precisión} \times \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}}$$

Figura 26: Fórmula para el cálculo de la medida-F en una búsqueda

Martínez Méndez recoge un amplio conjunto de medidas de evaluación de la recuperación de información (2004) y (2002). Estas medidas se aplican en el apartado 4.5. EVALUACIÓN DE LAS ONTOLOGÍAS ONA Y ONAPE para analizar la capacidad de la ontología para representar la información de una muestra de noticias.

*

**

3. PROCEDIMIENTOS METODOLÓGICOS

Los seis principios que se deben conseguir en la organización pueden resumirse así: más, mejor, más ordenado, más fácil, más rápido, para más usuarios (Otlet, 1934, p. 373bis)

3. PROCEDIMIENTOS METODOLÓGICOS

En este capítulo se describe la metodología que se ha seguido en el modelado y construcción de la ontología de dominio en política y economía, ONAPE, desde el momento de su concepción y hasta que queda implementada. Los procedimientos utilizados parten en buena medida de los principios metodológicos definidos por Mendonça y Almeida (2014), descritos en 2.2.3. METODOLOGÍAS PARA CONSTRUIR Y EVALUAR ONTOLOGÍAS. Por tanto, sigue varias de las cuestiones planteadas en metodologías reconocidas.

Antes se indican algunos aspectos metodológicos preliminares que, si bien no forman parte de la metodología propiamente dicha, sí son parte del proceso de aprendizaje desarrollado en esta tesis. También se especifica qué tipo de investigación se lleva a cabo.

3.1. ASPECTOS METODOLÓGICOS PREVIOS

Se relacionan las actividades formativas y laborales realizadas antes y durante el periodo de investigación que de alguna manera han influido en la misma:

- 1. Máster en Gestión de Información en las Organizaciones.** Para esta tesis es útil por lo que respecta a conocimientos básicos sobre investigación en Ciencias Sociales (bases de datos, referencias, estadística, elaboración de artículos científicos, etc.) y aspectos teóricos del área de Documentación, tales como indización, gestión de contenidos en la web y ontologías.
- 2. Cursos y congresos.** Se asistió y participó en cursos y congresos relacionados con Documentación periodística, investigación en Ciencias Sociales, web semántica, ontologías y metadatos. Se trata de temas en los que no se había profundizado lo suficiente anteriormente, de ahí la necesidad de este tipo de formación. La Tabla 1 recoge la lista de actividades.

Nombre del curso / congreso	Fecha	Institución
<i>Ayudante de Documentación de Medios de Comunicación</i>	2012	Servicio Regional de Empleo y Formación de la Región de Murcia; Formato Formación S.L.
<i>Cómo evaluar la calidad de la investigación en Ciencias Sociales</i>	2012	Centro de Formación y Desarrollo Profesional de la Universidad de Murcia (CFPD)

3. PROCEDIMIENTOS METODOLÓGICOS

Nombre del curso / congreso	Fecha	Institución
<i>Búsqueda eficaz</i>	2013	Fundación Germán Sánchez Ruipérez
<i>Introducción a la redacción de artículos científicos en lengua inglesa</i>	2013	Centro de Formación y Desarrollo Profesional de la Universidad de Murcia (CFPD)
<i>Aplicación de la Teoría de Grafos a la vida real</i>	2013	Universidad Politécnica de Valencia
<i>I Congresso ISKO-Espanha-Portugal. Informação e o Conhecimento: duas faces de Jano</i>	2013	Faculdade de Letras da Universidade de Porto (Portugal)
<i>International Conference on Dublin Core and Metadata Applications</i>	2013	ASSIST (Association for Information Science and Technology)
<i>Metadata: Organizing and discovering information</i>	2013	University of North Carolina at Chapel Hill
<i>Web semántica: estándares, herramientas y diseño de aplicaciones</i>	2013	SEDIC (Asociación Española de Documentación e Información)
<i>Redacción de artículos científicos en Ciencias Sociales</i>	2014	Centro de Formación y Desarrollo Profesional de la Universidad de Murcia
<i>Second Interdisciplinary Summer School on Ontological Analysis</i>	2014	IAOA (International Association for Ontology and its Applications)
<i>Software e-Termos, Ambiente Colaborativo web de Gestão Terminológica</i>	2014	Escola da Ciência da Informação de la Universidade Federal de Minas Gerais (UFMG)
<i>The landscape of information retrieval research. Research questions and methods</i>	2014	Escola da Ciência da Informação de la Universidade Federal de Minas Gerais (UFMG)
<i>Estadística para investigadores: Todo lo que siempre quiso saber</i>	2015	Universidad de Salamanca
<i>XII Congreso ISKO España y II Congreso ISKO España-Portugal</i>	2015	Facultad de Comunicación y Documentación de la Universidad de Murcia
<i>Elaboración de informes científicos y artículos</i>	2016	Centro de Formación y Desarrollo Profesional de la Universidad de Murcia (CFPF)
<i>Edición y Tratamiento de Imágenes (Gimp)</i>	2016	Centro de Formación y Desarrollo Profesional de la Universidad de Murcia (CFPF)

Tabla 1: Parte de cursos y congresos a los que se ha asistido

3. Estancias. Las estancias investigadoras han contribuido de manera especial al desarrollo de esta investigación. Gracias a éstas, se ha accedido a diversas bibliotecas, se han intercambiado ideas y opiniones con otros/as investigadores/as y los tutores han aportado sugerencias y mejoras para la investigación. Algunas estancias han dado lugar, además, a colaboraciones con otros/as investigadores/as. La Tabla 2 indica centro, universidad, fecha, duración y tutor/a de cada estancia.

3. PROCEDIMIENTOS METODOLÓGICOS

Centro	Universidad	Fecha	Duración de la estancia	Tutor/a
<i>Escola de Engenharia</i>	Universidade do Minho - Portugal	2013	4,5 meses	Ana Alice Baptista
<i>Escola de Ciência da Informação</i>	Universidade Federal de Minas Gerais (UFMG) - Brasil	2014	3 meses	Gercina Lima
<i>Faculdade de Filosofia e Ciências</i>	Universidade Federal Estadual Paulista Júlio Mesquita Filho (UNESP)	2016	1,25 meses	Edberto Ferneda

Tabla 2: Estancias durante el periodo de investigación

4. Idiomas. Todas las estancias han sido realizadas en países de habla portuguesa. Por este motivo, se han realizado cursos de este idioma y pruebas de nivel. Esto ha ampliado las posibilidades de acceso a literatura científica publicada en este idioma. Por otro lado, buena parte de la bibliografía está publicada en inglés, por lo que se han realizado cursos para mejorar los conocimientos de esta lengua. La Tabla 3 recoge algunos de los cursos y pruebas de nivel realizados durante el periodo de investigación.

Curso / Prueba	Idioma	Fecha	Centro	Nivel
<i>Português Língua Estrangeira</i>	Portugués	2013	Universidade do Minho	A2
<i>Inglés B1.2</i>	Inglés	2014	Universidad de Murcia	B1.2
<i>Diploma Elementar de Português Língua Estrangeira (DEPLE)</i>	Portugués	2015	Universidade de Lisboa	B1
<i>Formación en idiomas: inglés (PDI)</i>	Inglés	2016	Universidad de Murcia	B1+
<i>Formación en idiomas: inglés (PDI)</i>	Inglés	2016	Universidad de Murcia	B1.2
<i>Curso de Português para Estrangeiros</i>	Portugués	2016	Universidade Regional de Blumenau	---
<i>Língua Portuguesa</i>	Portugués	2016	Universidade Anhembi Morumbi	---
<i>Certificação de Proficiência em Língua Portuguesa para Estrangeiros</i>	Portugués	2016	Ministério da Educação (Brasil)	C1

Tabla 3: Cursos y pruebas de nivel de idiomas realizados durante el periodo de investigación

5. Experiencia laboral previa. Anteriormente al periodo de investigación se han tenido dos experiencias laborales: una beca de formación para personas diplomadas en Biblioteconomía y Documentación dedicada a la indización de noticias publicadas en prensa, y un contrato de documentalista para una empresa privada, que me permitió ser consciente de las limitaciones de la Web en cuanto a las posibilidades de recuperación de información y las dificultades por la falta de normalización en los procesos y en la descripción de documentos. Ambos puestos de trabajo se describen brevemente en la Tabla 4.

3. PROCEDIMIENTOS METODOLÓGICOS

Puesto de trabajo y lugar	Descripción
<i>Documentalista de prensa.</i> Oficina de Prensa de la Dirección General de Comunicación, Consejería de Presidencia de la Región de Murcia	1) Selección, digitalización y tratamiento documental de noticias de prensa regional y nacional en el gestor documental Invesdoc 2) Elaboración de Resúmenes de Prensa e Informativos y su correspondiente publicación en web
<i>Documentalista.</i> Global Ends, S.A.	1) Recuperación de información en línea 2) Creación y mantenimiento de archivos de oficina y definitivo 3) Normas ISO 9001:2000 y 9001:2008, etc.

Tabla 4: Puestos de trabajo anteriores relacionados con la investigación realizada

A continuación se indica qué tipos de investigación se han empleado para el desarrollo de esta memoria.

3.2. TIPO DE INVESTIGACIÓN

La investigación es la actividad que permite la aproximación y el entendimiento de la realidad, que pretende resolver un problema mediante procedimientos científicos (Silveira y Córdova, 2009, p. 31). Existen diversos tipos de investigación, que varían dependiendo del objeto de investigación: según su aproximación, su naturaleza (o grado de abstracción), sus objetivos y sus procedimientos. A continuación, se describe cada uno de los elementos anteriores para la investigación recogida en esta memoria

Por lo que respecta a la aproximación, es una investigación cualitativa, en la que se analizan las particularidades de la Gramática de Casos, de esquemas de metadatos para medios de comunicación escrita y de noticias de periódicos de ámbito internacional, para el modelado de acciones. Este tipo de investigación no se preocupa de aspectos numéricos, sino de la comprensión e interpretación de su objeto de estudio, sea un grupo social, una organización, etc. (Silveira y Córdova, 2009, pp. 31-33) o conjuntos de datos.

Atendiendo a la naturaleza o grado de abstracción de la investigación, se puede distinguir entre investigación básica y aplicada. La distinción entre ambas «debe ser entendida más como los extremos de un continuo que como una dicotomía. Ambos (...) convergen en su finalidad: hacer progresar el saber y simultáneamente transformar la realidad» (Martínez Mediano, 2014, pp. 35-36). El estudio pretende contribuir a la mejora de la recuperación de información respondiendo a problemas derivados del lenguaje natural y de la hiperinflación de contenidos. Por esto, puede hablarse de una investigación de naturaleza **aplicada**, ya que aplica el conocimiento con el propósito de resolver un problema, y se traduce en planes o directrices para su expansión (Martínez Mediano, 2014, p. 36; Silveira y Córdova, 2009, p. 35).

Teniendo en cuenta los objetivos planteados en la investigación, se puede hablar de tres tipos de investigación: exploratoria, descriptiva o explicativa, o una combinación de las anteriores. En este caso, es una investigación de tipo exploratorio-descriptivo. La

3. PROCEDIMIENTOS METODOLÓGICOS

investigación exploratoria «tiene como objetivo proporcionar mayor familiaridad con el problema, con el fin de hacerlo más explícito o construir una hipótesis» (Silveira y Córdova, 2009, p. 35) e implica entre sus procedimientos el análisis de bibliografía. La investigación descriptiva busca «describir el comportamiento de un determinado acontecimiento, sujeto o grupo de sujetos» (Martínez Mediano, 2014, p. 37) y se propone básicamente medir una o más variables en alguna población o muestra (Gutiérrez-Cillán y Rodríguez Escudero, 2013, p. 51). Además, se caracteriza por que el objeto de análisis, en este caso, acciones del mundo real recogidas en noticias, sólo es observado y analizado para describir hechos, sin alterar nada (Martínez Mediano, 2014, p. 49). Esto es, análisis documental (Silveira y Córdova, 2009, p. 35).

En relación con las técnicas para la recogida de datos, se pueden emplear varias, entre las que Silveira y Córdova (2009, pp. 36-42) distinguen la investigación experimental, bibliográfica, documental, de campo, *ex-post-facto* (causa-efecto), de análisis, mediante entrevistas, estudios de caso, investigación participante, investigación-acción, etnográfica o etnometodológica. Gil Pascual (2011, p. 23), por su parte, indica las siguientes: «observación, entrevista, cuestionario, pruebas objetivas y tests, técnicas grupales y sociométricas y análisis de documentos», indicando posteriormente (2011, p. 281) que el análisis de documentos es «el fundamento de toda investigación que pretenda estar incluida en la realidad del momento en el que se desarrolla». En esta investigación se aplican técnicas bibliográficas, técnicas documentales y observación.

Las técnicas bibliográficas son empleadas en la fundamentación teórica y la descripción de vocabularios, así como esquemas de metadatos utilizados en periódicos, con el propósito de conocer qué se ha estudiado sobre el asunto de análisis. Las técnicas documentales se emplean sobre la muestra de noticias, analizando cómo se estructuran, qué y cómo narran acontecimientos del mundo real, mediante indización, humana y automática. La diferencia entre técnicas bibliográficas y documentales tiene que ver con el tratamiento de la fuente. En cuanto a la observación, Gil Pascual (2011, pp. 107-125) señala los aspectos que intervienen: el/la observador/a, «que codifica las situaciones o eventos ocurrentes; el instrumento de registro (...) y la situación observada, que es un complejo de múltiples eventos y relaciones, del cual hay que seleccionar los elementos a estudiar». En relación con los instrumentos de registro, se indican las fuentes de obtención de información y cómo se han analizado. Sobre la situación observada, se tienen en cuenta dos realidades: 1) científica: artículos científicos, vocabularios, esquemas de metadatos; 2) social: acontecimientos del dominio político-económico, a partir de su registro en noticias de medios de comunicación impresa y su análisis mediante indización.

En el siguiente apartado se pone el foco de atención en la metodología propiamente dicha.

3.3. LIMITACIONES DE LA INVESTIGACIÓN

Durante la investigación se han identificado una serie de limitaciones. En primer lugar, destaca la brecha de conocimiento relacionada con las ontologías y su modelado. Mientras esta área es tratada de forma habitual en los estudios relacionados con Informática, no ocurre igual en el área de Documentación, o es tratado sucintamente. El problema para el/la documentalista es «que deberá conocer y manejar una serie de conceptos nuevos propios de la lógica (lógica de predicados, descriptiva, etc.), la matemática discreta (propiedades de relaciones: reflexiva, simétrica, transitiva, etc.), la ingeniería del conocimiento (técnicas de educación y formalismos de representación del conocimiento, etc.), aparte de aquellos propios de su profesión, como los relacionados con las bases de datos (formatos de representación de la información: números reales, enteros, booleanos, etc.)» (García-Torres et al., 2008).

Otra limitación, inherente a todo análisis humano, es la subjetividad en que se puede incurrir en algunos procesos en esta investigación: la indización humana, la descripción de noticias e, incluso, en la construcción de la ontología. Ejemplos de esta cuestión tienen que ver con el dominio analizado, del tipo ¿cómo clasificar “aborto”? ¿Es un derecho de la mujer o un asesinato? ¿Cómo categorizar aspectos relacionados con la “ocupación de un territorio” o “invasión de un territorio”? ¿Qué punto de vista tomar? ¿Puede estar un territorio legítimamente ocupado? ¿A partir de qué criterios se considera como tal? ¿Qué se debe tener en cuenta, lo que dice un periódico, lo que dice otro, una fuente oficial? ¿Cuál? ¿De qué país? Por otro lado, aún intentando tener una visión objetiva, es imposible eliminar totalmente cierto sesgo personal, incluso la influencia eurocentrista, por ejemplo, puede llevar a utilizar cierta categorización en lugar de otra.

Una tercera limitación tiene que ver con las condiciones de acceso al texto completo de las noticias. Además de la imposibilidad de pago por el acceso a los periódicos de la muestra, puede darse la circunstancia de que un medio gratuito sea de pago en cualquier momento. En este sentido, titulares (y metadatos en el código fuente de las noticias) son siempre accesibles. En relación con la caducidad de las noticias en Internet (URL temporales), se solucionó almacenando las noticias en Evernote y, posteriormente, en MetadadosHTML.

También la diversidad de lenguas de los diarios de la muestra dificulta el tratamiento de la información. Esta variabilidad, por otro lado, aporta riqueza y puntos de vista diferentes.

Finalmente, la indización humana, pero también la automática, puede verse comprometida por el tamaño de dicha muestra, de ahí la necesidad de reducir los *corpora* a la submuestra₁. Aun así, el volumen de datos a analizar es grande para revisarlo en un tiempo razonable. Por ello, en la evaluación de ONA y ONAPE se opta por analizar casos concretos, con la intención de seguir poblando la ontología en trabajos futuros.

3.4. DESARROLLO DE LA INVESTIGACIÓN

La metodología que se plantea para el modelado y construcción de una ontología de dominio tiene como base los principios metodológicos recogidos por Mendonça y Almeida (2014), resumidos en 2.2.3. METODOLOGÍAS PARA CONSTRUIR Y EVALUAR ONTOLOGÍAS.

Como resultado, se obtienen dos ontologías: ONtología de Acciones (ONA) y ONtología de Acciones de Política y Economía (ONAPE). Ambos modelos recogen algunas instancias, aunque sólo a modo de ejemplo, por lo que, en puridad, debe hablarse más bien de un conjunto de elementos de ontología. Las etapas seguidas, que no siempre siguen una línea temporal y pueden solaparse unas con otras, son resumidas a continuación.

CONFIGURACIÓN DE LA MUESTRA

Para la construcción de ONAPE se recurre a términos que son utilizados realmente más que los que deberían ser utilizados, lo que resulta más útil que otras aproximaciones (Stuart, 2016, p. 98). Esta idea también es sostenida por Hlomani (2014), que destaca la importancia de elegir una fuente válida para la extracción de los términos que formarán parte de la ontología. En este sentido, para la configuración de la muestra se llevó a cabo una selección continua de zonas geopolíticas, países y periódicos, mediante un muestreo no aleatorio intencional. Los datos de los medios de comunicación escritos se extrajeron de la web 4 International Media and Newspapers (4IMN). Diariamente, durante un año, para cada medio se seleccionó manualmente una noticia, de acuerdo a criterios definidos, que se indican en la sección correspondiente.

Para el almacenamiento de los datos se recurrió a la extensión Clearly, de Evernote, disponible en Google Chrome. Ésta permite almacenar el texto completo cuando es posible, titular, fecha de publicación y url. Después, se descargaron todas las noticias en formato enex y se convirtieron a csv para facilitar su tratamiento.

Del conjunto de la muestra, se definieron submuestras tanto para la indización como para la evaluación de resultados:

La submuestra₁ es explotada en los procesos de indización descritos en 2.3.1. INDIZACIÓN HUMANA E INDIZACIÓN AUTOMÁTICA, aprovechando el metadato título extraído con Evernote. El empleo de titulares se justifica por:

1. La limitación del acceso a noticias como solución a problemas de rentabilidad de los medios de comunicación impresa. Se están estableciendo *paywalls* que impiden el acceso a parte de la información, de actualidad o retrospectiva (Hagey, 2012);
2. Están siempre accesibles. Los titulares son datos casi siempre disponibles, aunque los periódicos establezcan muros de pago;

3. PROCEDIMIENTOS METODOLÓGICOS

3. La indización es más simple cuanto menor sea el contenido a analizar. Además, el titular recoge buena parte de los elementos representativos del contenido de la noticia. Los titulares son un elemento básico de recogida de los elementos principales de una noticia (ver 2.3.3. INDIZACIÓN DE NOTICIAS. LA GRAMÁTICA DE CASOS). Estudios previos ya han empleado los titulares como fuente para el modelado de ontologías. Por ejemplo, la ontología *News Ontology Event Model* (NOEM), de Wang y Zhao (2012), también parte de un análisis de titulares de noticias, de las agencias XinHua y People. Los autores aseguran que el 85% de las noticias pueden ser descritas con su ontología.

La submuestra₂ está compuesta por 23 noticias distribuidas en cinco temas del dominio de ONAPE, política-economía. Este grupo de noticias es utilizado para la validación de ONAPE (y ONA) mediante instanciación en dos etapas. Para la validación basada en instanciación de palabras clave del código fuente, se recurre al software MetadadosHTML. Las características de esta submuestra se describen con mayor detalle en 4.5.1. PREPROCESAMIENTO DE LA SUBMUESTRA₂ PARA LA EVALUACIÓN DE ONAPE: METADADOSHTML. Para la validación basada en instanciación del texto completo se utilizan las noticias del tema 1 de esta submuestra. El texto completo de estas noticias, marcadas con ONA, está disponible en el anexo C. MARCADO SEMÁNTICO DE NOTICIAS DEL TEMA 1.

MODELADO DE LA ONTOLOGÍA DE ACCIONES ONA BASADO EN LA GRAMÁTICA DE CASOS

Partiendo del marco teórico de la Gramática de Casos (Fillmore, 1967; García Gutiérrez, 2014), se plantea un modelo inicial de ontología general para la descripción de acciones, ONtología de Acciones (ONA).

Para ello, se analizan los componentes de esta teoría y se establecen relaciones entre estos. En un primer planteamiento, cada elemento del modelo teórico es convertido en una clase, relacionada con otras mediante propiedades tipo objeto.

En un segundo análisis, descrito en 4.2.2. MODELADO INICIAL DE ONA, se tienen en cuenta las particularidades de la formalización del modelo en OWL, por lo que se reducen las clases indicadas inicialmente. Por esta cuestión, determinados aspectos vinculados a las acciones como, por ejemplo, su causalidad, se expresan mediante propiedades y no por clases, que quedan definidas de forma más general. Además, se incorporan relaciones entre las clases especificadas y dos clases intermedias creadas ad hoc, Sujeto y Objeto. Éstas facilitan la identificación de la participación conjunta de dos o más elementos en una acción, sea como ejecutante o como receptor de dicha acción.

3. PROCEDIMIENTOS METODOLÓGICOS

Con vistas a la formalización de ONA en Protégé, se siguen las recomendaciones de Aguado-de-Cea et al. (2015) y Horridge (2011). Aquí pues, ya se definen algunas cuestiones de formalización de las ontologías. Por ejemplo, cómo citar las fuentes de definición de un concepto o qué estilo de anotación adopta la etiqueta. Dichas recomendaciones se especifican en 2.2.1. COMPONENTES DE LA ONTOLOGÍA.

En esta etapa se ha definido buena parte de lo que será ONA, pero aún no es un producto acabado; en procesos posteriores es modificada.

DESCUBRIMIENTO DE CONOCIMIENTO POR INDIZACIÓN. MODELADO DE LAS ONTOLOGÍAS

Teniendo en cuenta el modelo inicial ONA y, con vistas a la incorporación de subclases y subpropiedades, se indizan las noticias de la submuestra₁, determinando la temática de la noticia y su traducción a un conjunto de términos (Lancaster, 2004, p. 9). Se emplean procesos tanto de indización humana como automática. Cada término atribuido a las noticias o extraído por e-Termos es, potencialmente, un elemento a recoger por ONA, en caso de términos generales, y por ONAPE (ONtología de Acciones de Política y Economía) en caso de términos especializados.

La extracción de términos por indización automática es asistida por el software e-Termos (*Termos Electrônicos*), descrito en 2.3.2. EXTRACCIÓN DE TÉRMINOS CON N-GRAMAS: e-TERMOS. ANÁLISIS MORFOLÓGICO: GRAMPAL Este software aplica la técnica n-gramas, con combinaciones de una a siete gramas. Primero se preparan los *corpora* que van a ser sometidos a indización automática, traduciendo los titulares a un único idioma, español, y suprimiendo aquellos elementos propios de éste que pueden generar problemas (eñes y tildes). Además, se asigna una lista de palabras vacías al proceso. Después, se especifica el tipo de documento con el que e-Termos va a trabajar. Finalmente, se obtienen datos (listas de términos) para cada grama, desde 1 grama (un término) a 7 (términos compuestos por 7 palabras). En el caso de unigramas, los términos obtenidos fueron sometidos a un proceso de lematización con el software Grampal.

Los términos atribuidos en la indización humana pueden entenderse como conceptos, ya que son asignados teniendo en cuenta los fenómenos propios del lenguaje natural, como género y número, sinonimia, polisemia, etc., y evitando errores tipográficos. Por el contrario, los elementos extraídos en e-Termos son analizados para reducir el ruido (errores de e-Termos). Cada lista es analizada teniendo en cuenta las mismas cuestiones propias del lenguaje natural con que se realizó la indización humana y se crea una única lista de elementos potenciales a incluir en las ontologías.

Posteriormente, tanto para la lista de términos asignados por indización humana como para la lista de términos extraídos por indización automática, cada elemento es analizado

3. PROCEDIMIENTOS METODOLÓGICOS

individualmente. En cada caso, se valora la pertinencia de su incorporación al vocabulario, tanto en ONA como en ONAPE. Como en todo proceso humano, la decisión de lo que es o no es una clase, subclase o propiedad no es absoluta, y diferentes personas pueden tomar decisiones distintas para diversas ontologías (Stuart, 2016, p. 31).

Para el modelado de ONAPE se adopta el enfoque *middle-out*, que intercala los enfoques *top-down*, *bottom-up* y favorece el descubrimiento de conceptos durante la construcción de la ontología (Beydoun et al., 2011). Este procedimiento consiste en definir los términos fundamentales y, posteriormente, ir definiendo términos más abstractos y más específicos. Por ejemplo, “Mamífero” es una generalización de “Perro” y “Cocker spaniel” es una especialización (Uschold y Grüninger, 1996, pp. 20-22). Para estos autores, este enfoque facilita la relación entre términos en el área de forma más precisa y reduce la duplicidad del trabajo. De esta forma, para la definición de conceptos y relaciones, así como para su clasificación en ambos vocabularios, se recurre a varias fuentes, como el Diccionario de la Real Academia de la Lengua Española (DLE), Wikipedia, los tesauros EUROVOC, UNBIS Y AGROVOC y varios de los vocabularios de los que se plantea su mapeado con ONA y ONAPE, como schema.org, FOAF o TIME. Algunas de estas herramientas son descritas en 4.4.1. PRESELECCIÓN DE VOCABULARIOS y A.1. HERRAMIENTAS REUTILIZADAS POR LOS VOCABULARIOS. En este análisis, en el que además se tienen en cuenta las definiciones de los términos obtenidos por indización, se aplican los dos métodos comunes para la creación de vocabularios controlados: inducción, por el que se agregan elementos al vocabulario a partir de la indización de documentos (Keyser, 2012, p. 25); y deducción, por el que se recogen elementos que no son fruto expreso de la indización, sino de sus definiciones y relaciones con otros elementos en las herramientas indicadas. En la formalización de ONAPE también se siguen las recomendaciones de Aguado-de-Cea et al. (2015) y Horridge (2011).

REUTILIZACIÓN DE ELEMENTOS PREVIOS: MAPEO CON OTROS VOCABULARIOS

Se analizan diversos vocabularios previos, como FOAF, TIME, schema.org y PRoles para su reutilización en ONA y ONAPE, en el caso de vocabularios de dominio.

Como fuente para la selección de estos vocabularios se utiliza el repositorio *Linked Open Vocabularies* (LOV), descrito en 2.2.5. REPOSITARIOS DE VOCABULARIOS: LINKED OPEN VOCABULARIES, y bases de datos de artículos. Siguiendo como criterios la adecuación a la Gramática de Casos, la difusión, el tamaño y el nivel de abstracción, se preseleccionan dieciséis vocabularios y se profundiza en su análisis. Después, se alinean con el modelo, indicando qué clases y relaciones podrían importarse.

3. PROCEDIMIENTOS METODOLÓGICOS

Este análisis pone en evidencia que ninguna de las ontologías analizadas cubre totalmente el dominio ni de ONA ni de ONAPE, aunque sí pueden reutilizarse algunos de sus elementos. Sin embargo, no se ha incorporado ninguno, como podría ser, por ejemplo, la clase *Agent* de FOAF. Es necesario un análisis en mayor profundidad de los vocabularios con que mapear. Por esta cuestión, se ha previsto como investigación en el apartado 5.2. TRABAJOS FUTUROS. Un punto de partida para dicho estudio, en ambas ontologías, es la propiedad *rdfs:isDefinedBy*, que se ha empleado para especificar la fuente de definición de la clase o propiedad. De esta forma, en una versión posterior de los vocabularios, si así se establece, estos elementos podrían ser importados y no referenciados como fuente.

El apartado 4.4. REUTILIZACIÓN DE ELEMENTOS PREVIOS: MAPEO CON OTROS VOCABULARIOS describe cómo se plantea el mapeo de ONA y ONAPE con los vocabularios seleccionados para dicho proceso. La extensión de la descripción de estas herramientas justifica su ubicación en esta memoria como anexo: A. DESCRIPCIÓN DE VOCABULARIOS MAPEADOS.

EVALUACIÓN DE LAS ONTOLOGÍAS ONA Y ONAPE

La penúltima fase de modelado y construcción de ONAPE (y ONA) es su evaluación, tal como se indica en 2.2.3. METODOLOGÍAS PARA CONSTRUIR Y EVALUAR ONTOLOGÍAS. Esta evaluación comprende dos procesos: verificación y validación.

La verificación es efectuada automáticamente en Protégé, mediante el razonador Hermit, aplicado tanto a ONA como a ONAPE. La validación comprende el análisis de la adecuación de ONAPE (y con ella, de ONA) al propósito para el que se crea. Se plantea una validación inicial en dos etapas:

- a) **Poblamiento de ONAPE** a partir de las palabras clave extraídas con el software MetadadosHTML del código fuente de las noticias de la submuestra₂. Las palabras clave forman parte del conjunto de metadatos que describen y facilitan el almacenamiento y recuperación (Troncy, 2008) de piezas de información. Existen múltiples esquemas de metadatos para artículos de prensa. Por ello, previamente a la validación, se ha analizado qué esquemas de metadatos utilizan y qué prácticas de descripción siguen los periódicos de la muestra, lo que da lugar a distintos grados de normalización, tanto en el uso de los esquemas como en la inclusión de palabras clave. La toma de datos para este análisis (tres noticias seleccionadas al azar por diario), se realizó en diciembre de 2014, febrero y abril de 2015. A partir de esta revisión se establecen una serie de criterios para la extracción de palabras clave del código fuente de noticias de la muestra y se construye MetadadosHTML.
- b) **Marcado semántico** de texto completo de las noticias del temas 1 de la submuestra₂.

3. PROCEDIMIENTOS METODOLÓGICOS

Para acabar esta etapa, se aplica una adaptación de tres medidas clásicas de análisis de la indización: exhaustividad, precisión, así como una combinación de ambas, la medida-F. Estas fórmulas se describen brevemente en 2.3.4. EVALUACIÓN DE LA INDIZACIÓN. Exhaustividad y precisión son destacadas por Hlomani (2014) como medidas de evaluación de la exactitud de una ontología. Para el cálculo de estos índices, en lugar de documentos de un sistema y documentos recuperados, se tienen en cuenta términos de documentos y términos de ONAPE.

Este proceso de validación puede ser entendido como paso inicial del aprendizaje que requiere la máquina para, en un futuro, ser capaz de obtener nuevas instancias de la ontología (instanciación o poblamiento) así como de describir conjuntos de noticias. Por tanto, se plantea una investigación a partir de 5.2. TRABAJOS FUTUROS.

La última etapa es 4.6. ESPECIFICACIONES DE LAS ONTOLOGÍAS, donde se detalla por qué surge dicho vocabulario y sus distintos elementos. Se utiliza LODE (*Live OWL Documentation Environment*) para generar automáticamente esta documentación. Únicamente es necesaria la URL de las ontologías o bien su fichero, en OWL o RDF.

3.5. APLICACIONES Y PÚBLICO OBJETIVO

ONA y ONAPE, como vocabularios de representación del conocimiento, pueden ser útiles en diversos aspectos:

1. **Clasificación de documentos.** Algunas clases de ONAPE pueden ser empleadas para la clasificación de documentos de un sistema de información, facilitando un primer acercamiento a sus contenidos. Un ejemplo de aplicación de esta posibilidad es la clasificación por tipología documental de la Figura 27.



Figura 27: Tipologías documentales contempladas por ONAPE

3. PROCEDIMIENTOS METODOLÓGICOS

Otra posibilidad es la clasificación de documentos por sector de actuación, como *SectorEducativo* o *SectorComunicación*.

- 2. Navegación semántica basada en autoridades.** Un aspecto relacionado con el anterior es el marcado semántico como elemento para el descubrimiento de contenidos por navegación. Esto sucede especialmente en los medios de comunicación, ya convertidos en buena parte en cibermedios. Se necesitan «palabras clave para que los diferentes sistemas de navegación del sitio web proporcionen el mejor acceso posible a noticias de una misma temática» (Codina y Pedraza-Jiménez, 2017). A partir de las entidades básicas de ONA o de elementos más específicos de ONAPE, se accede a los contenidos que han sido marcados con dichos elementos. Estos contenidos están relacionados con otros tantos mediante las distintas propiedades de la ontología, ya sean, por ejemplo, ejecución de acciones, causas o consecuencias. Cada vez que se accede a un contenido, aparecen distintos enlaces que llevan a otros elementos (instancias de ONAPE) con el que mantienen algún vínculo. Esta cuestión resulta especialmente práctica en aquellos sistemas que manejan una gran cantidad de información y/o que necesitan que sus usuarios/as puedan moverse entre sus documentos (sus metadatos) no sólo jerárquicamente, sino horizontalmente. Como destaca Codina (2017), la estructura de navegación de un sitio contempla dos grandes tipos de navegación: estructural y semántica. Añadir un vocabulario a la navegación semántica facilitaría dicha navegación e impediría el crecimiento desmesurado de términos de indización.

Una forma de identificar el potencial de este aspecto es el portal <http://datos.bne.es/inicio.html>, desarrollado por el Grupo de Ingeniería de Ontologías de la Universidad Politécnica de Madrid y publicado en enero de 2017.

La ontología, frente a otros vocabularios, tendría como ventaja la posibilidad de establecer relaciones entre las clases. De esta forma, el sistema permitiría una búsqueda mucho más rica, gracias a las opciones de navegación entre los distintos elementos semánticos atribuidos a los documentos. El funcionamiento con ONAPE sería similar al de aquellos sistemas que relacionan conceptos y documentos mediante relaciones asociativas, aunque con un mayor nivel de detalle. Una ontología, de hecho, «podría servir para proporcionar un sistema de búsqueda inteligente, capaz de realizar inferencias y actuar como si fuera una verdadera inteligencia artificial, pero aún no están implantadas en forma funcional» (Codina y Pedraza-Jiménez, 2017).

3. PROCEDIMIENTOS METODOLÓGICOS

3. Sistema de representación y recuperación de los contenidos en una organización del dominio analizado, basado en la ontología modelada. Se trata de un sistema más complejo, en el que ONAPE es empleado tanto en la fase de marcado semántico (o indización) como en la recuperación de documentos mediante asistentes que orienten al usuario/a en el empleo de los elementos contemplados por el modelo, como se representa en la Figura 28.

Los resultados de una búsqueda pueden mejorar si, además, se combinan descripción de documentos y ecuación de búsqueda (empleando la ontología) con la experiencia de sus usuario/as. Una posibilidad es la atribución de mayor o menor relevancia de los documentos recuperados respecto de la ecuación de búsqueda, como ya apuntaban Salton y McGill (1983, pp. 79-83). De esta forma, el *ranking* de documentos mostrado tras una búsqueda cambia con cada búsqueda. Es lo que el autor denomina valor dinámico de la información (*dynamic information value*).

Cualquier elemento:	<input type="text"/>		
Acción:	<input type="text"/>	Autoridades	Y/O/NO
Agente:	<input type="text"/>	Autoridades	Y/O/NO
Cosa:	<input type="text"/>	Autoridades	Y/O/NO
Espacio físico:	<input type="text"/>	Índice geográfico	Y/O/NO
Espacio Temporal:	<input type="text"/>	Índice temporal	
		Buscar	Ordenar por

Figura 28: Hipotético sistema de recuperación de información basado en ONAPE

La adición de Wordnet permitiría, además, no sólo introducir una granularidad más fina en la búsqueda, sino la expansión de preguntas de investigación (Rocha, 2012). Wordnet es, sin duda, el recurso externo más utilizado (Stuart, 2016, p. 120). La traducción de las entidades a otras lenguas ampliaría las posibilidades de la indización, tanto humana como automática. Esto ya ha sido utilizado en otros trabajos como Fernández-García (2006) o Ruiz-Martínez, Baños-Moreno y Martínez-Béjar (2014). Este aspecto es contemplado como futura investigación en el apartado 5.2. TRABAJOS FUTUROS.

Otra opción es el uso de ONAPE en sistemas de aprendizaje automático. Estos sistemas están preparados para que a medida que manualmente se instancia una ontología y se describen textos con ésta, estos aprendan a identificar patrones que

3. PROCEDIMIENTOS METODOLÓGICOS

se repiten en dichos textos. En este sentido, la validación de la ontología, que se recoge en el apartado 4.5.4. VALIDACIÓN POR ANOTACIÓN SEMÁNTICA puede entenderse como un punto de partida.

4. **Gold standard:** en los proyectos en los que se aplica el PLN es habitual el uso de gold standards. Estos son definidos como «conjunto de términos de referencia con todas las formas aceptadas [en el proceso]» (Moreno Sandoval y Campillos Llanos, 2015). ONAPE ha sido validada en dos ocasiones (ver 4.5. EVALUACIÓN DE LAS ONTOLOGÍAS ONA Y ONAPE) y aunque requiere de un desarrollo mayor y su instanciación, sus términos puede ser empleados como *gold standard* para el dominio política-economía.
5. **Esquema de metadatos de propósito general, basada en ONA, o de propósito específico, basado en ONAPE:** otra forma de marcar el contenido de los documentos es a partir de esquemas de metadatos. Un esquema basado en cualquiera de los dos vocabularios facilitaría la interoperabilidad con sistemas que emplean estas herramientas, así como la incorporación de elementos de interés en otros esquemas, como por ejemplo, Instrumento, a través de perfiles de aplicación. Los esquemas de metadatos han cobrado renovada importancia con la llegada de la administración electrónica y la necesidad de mejorar la descripción de documentos para una recuperación de información precisa. En determinados contextos, como el judicial, ONA podría ser útil para identificar quién comete un delito, cuál era su intención, causas, consecuencias, etc.

Por lo que respecta al público objetivo, ONAPE puede ser útil para diferentes sectores de la población, entre ellos periodistas, personas dedicadas a la política, economía y sociología, e historiadores, documentalistas y ciudadanía en general, ya que contribuye a la representación de contenidos de documentos de un dominio de interés general en un contexto concreto (medios de comunicación impresa), así como a su recuperación mediante la implementación del modelo en un sistema de recuperación de información.

La utilidad para cada uno de los perfiles indicados se puede resumir de la siguiente forma:

- **Periodistas:** acceso a documentos almacenados en un sistema, sea el propio medio en el que trabajan o en otros. Esto facilita una mejor descripción y comprensión de los hechos acontecidos, lo que se puede traducir en una mejor contextualización de su trabajo. Las exigencias de rapidez en la redacción y publicación de artículos, además, pueden verse comprometidas por el ruido y el silencio en la recuperación de información. Esto se debe a la ausencia de herramientas de control e indización.

3. PROCEDIMIENTOS METODOLÓGICOS

- **Personas dedicadas a la política o economía:** como parte de su trabajo, los profesionales del dominio de análisis deben estar informados de lo que ocurre o ha ocurrido en su ámbito de actuación. Documentos recuperados (a través de procesos de indización y consulta basados en ontologías) pueden contribuir a una mejor toma de decisiones.
- **Personas dedicadas a la sociología e historia:** política-economía son ámbitos ampliamente estudiados (análisis del pasado de los pueblos y de la sociedad). Como en los casos anteriores, una mejor descripción de los documentos especializados facilita su recuperación y un mejor aprovechamiento del tiempo, que puede ser empleado en otras actividades como la redacción y publicación de trabajos científicos.
- **Documentalistas y profesionales afines:** como intermediarios en la descripción de documentos y su puesta a disposición para los usuarios de un sistema de información, el modelo es útil para la mejora de los resultados de interfaces en sistemas informáticos de recuperación de información. El modelo, además, desde el punto de vista de la investigación en Documentación e Informática, puede ser utilizado para construir una ontología de acciones en otros dominios, utilizando para ello los elementos más generales de dicho modelo.
- **Ciudadanía, en general:** para este público, el modelo puede contribuir con la descripción y recuperación de documentos, especialmente noticias, que faciliten el conocimiento de acciones en el dominio y la toma de decisiones basada en estos. El caso más evidente, por ejemplo, sería decidir a qué partido político votar en unas elecciones en función de las noticias publicadas sobre sus acciones, causas, consecuencias, intenciones, etc. Asimismo, un uso concreto hacia el que podría dirigirse el modelo es la oferta personalizada de noticias con un contenido muy específico. De esta forma, el modelo contribuiría a «cubrir el hueco que existe entre la necesidad de los usuarios para la selección personalizada de contenidos y lo que la industria mediática puede ofrecer» (Fernández-García et al., 2007).

*

**

4. RESULTADOS

En los últimos años han aparecido nuevos términos en el campo de la Documentación, como (...) ontologías. En realidad, estos términos son usualmente aplicados a conceptos más viejos y conocidos, [...] ¿por qué estos cambios de significado? La causa es la entrada de las tecnologías de la información en el mundo de la Documentación e Información [...]. Estos cambios han ocurrido y hoy las ontologías son sinónimo de lenguajes documentales, sólo que tienen estructuras bastante diferentes (Currás, 2010, pp. 18-19).

Se describen los procedimientos metodológicos, así como los resultados de su ejecución. Aunque en este apartado se muestran estos procedimientos de forma secuencial, en ocasiones estos se superponen, desarrollándose simultáneamente. Este capítulo comienza indicando cómo se ha construido la muestra de noticias a partir de la que obtener una fuente de términos con la que construir el modelo ONAPE y evaluarlo.

4.1. CONFIGURACIÓN DE LA MUESTRA

En el modelado y construcción de una ontología, la utilización de documentos relativos al dominio que se modela es un buen descriptor de dicho dominio (Hlomani, 2014, p. 40), de ahí la importancia de conocer dichos modelos. En el siguiente apartado se indican algunas características de la noticia como documento objeto de indización. Después se detalla cómo se ha construido la muestra de documentos, mediante una selección continua de zonas geopolíticas, países y periódicos, siguiendo un proceso de muestreo no aleatorio intencional. Para los datos de los medios se recurrió a la página 4 *International Media and Newspapers* (en adelante, 4IMN).

4.1.1. LA NOTICIA, FUENTE DE TÉRMINOS PARA LA REPRESENTACIÓN DEL CONOCIMIENTO

Una noticia es un texto que relata un hecho susceptible de ser de interés público, por ser algo novedoso, destacable y/o actual. Es el género informativo por excelencia (García Gutiérrez, 2014) y el más habitual en los medios de comunicación escrita. El dominio política y economía es ampliamente tratado en las noticias de periódicos especializados, pero también en los generalistas (Quandt, 2008). Los periódicos suelen dar cobertura a noticias tanto de ámbito nacional como internacional, con temas muy distintos, como creación de empresas, actualización de datos sobre las bolsas de distintos países, conflictos y casos de corrupción, nombramientos de cargos públicos etc. La selección manual de noticias de la muestra asegura su adecuación al dominio de análisis.

Entre las características de la noticia destacan su actualidad e inmediatez, la redundancia **cíclica** (las mismas categorías y temas entran y salen de la *agenda setting* (Allen, Japzon, Achananuparp y Lee, 2007; García et al., 2006; Liddy, McVeary, Paik, Yu y McKenna, 1993)) y coocurrencia (relato de un mismo hecho por medios diferentes). Estos dos últimos

4. RESULTADOS

aspectos aportan como ventaja para la indización la posibilidad de incrementar la resolución semántica de una acción, «constituida por la pluralidad y diversidad de noticias sobre el evento» (Fidalgo, 2004, p. 61). De esta forma, se facilita el acceso a los conceptos y términos más novedosos y actuales de un dominio, el establecimiento de conexiones temporales con otros momentos, el análisis desde diversas perspectivas y la identificación de contenidos por repetición. Como consecuencia, se reduce la incertidumbre de determinar qué unidades léxicas son relevantes, pues serán destacadas por varios medios en el espacio y/o en el tiempo.

Otra cuestión en cuanto a la noticia como fuente de obtención de términos y conceptos es el tamaño del documento. Éste suele tener una extensión menor, en general, que cualquier otro documento de trabajo, lo que agiliza el análisis de contenido. También aporta cierto contraste de la información antes de su publicación, debido a las políticas de verificación de lo acontecido por parte de los medios, descartando además lo que se considera de menor interés (Río Cabrerizo, 2014, p. 69). Esas políticas también contemplan guías de estilo sobre cómo transcribir determinados neologismos, extranjerismos, etc. De esta forma, la noticia puede contribuir a la normalización de la terminología.

Por otro lado, las noticias de dominios especializados, como política y economía, suelen ser redactadas por especialistas o periodistas con conocimientos en el área, que interpretan la información de acontecimientos complejos y los “traducen” para facilitar la comprensión de la acción por los lectores del medio.

Finalmente, existe cierta categorización de la información en el propio medio, con las posibilidades que se muestran en la Tabla 1. Muchos de estos elementos de categorización aparecen con frecuencia en los titulares, pues, como se indica en 2.3.3. INDIZACIÓN DE NOTICIAS. LA GRAMÁTICA DE CASOS, suelen responder a las preguntas de qué ocurre, quién ejecuta esa acción, cuándo ocurre y dónde tiene lugar.

Elemento de categorización	Descripción del elemento
<i>Secciones</i>	Habitualmente los artículos se organizan siguiendo esquemas similares a los diarios en papel, por secciones tales como: nacional, política, economía, empresas, sucesos, etc.
<i>Tópicos</i>	Para acontecimientos destacados, los medios suelen crear etiquetas con las que agruparlos para facilitar el acceso a noticias relacionadas
<i>Materias</i>	Actúan de forma similar a los tópicos, aunque son más genéricas y no vinculadas a instancias específicas. Una diferencia entre tópicos y materias sería: elecciones generales 2016 (tópico) y elecciones (materia)
<i>Entidades nombradas</i>	Descripción del contenido de los artículos a partir de elementos vinculados a una herramienta de gestión del conocimiento. Pinchando en estos elementos se podría acceder no sólo a noticias sobre esos elementos, también a otros aspectos relacionados

Tabla 1: Elementos de categorización de noticias

4. RESULTADOS

De hecho, para autores como Salton (1975, p. 86), los términos que aparecen en títulos de documentos podrían ser utilizados en exclusiva para representar el contenido de dichos documentos o, al menos, se les podrían dar un mayor peso. Sin embargo, García Gutiérrez (2014) destaca que «los títulos de las noticias no siempre proporcionan toda la información relevante (...) e incluso suelen responder a criterios comerciales, políticos, estilísticos o de índole no netamente textual». La elección de titulares descriptivos podría solventar esta cuestión.

Como inconveniente, los conceptos en el área de las ciencias sociales y humanidades, como es el caso de las noticias de política-economía, se caracterizan, entre otros, por cuestiones como la ambigüedad, subjetividad, parcialidad, etc. (Raber, 2003, p. 132; Ruíz-Martínez, 2011, p. 131), que pueden afectar a la indización, pero también a la interpretación de los hechos narrados en el propio documento. En el ámbito periodístico surgen además otros problemas, como por ejemplo, que el periodista presuponga por parte del lector el conocimiento del hecho narrado, obviando información relevante.

4.1.2. SELECCIÓN DE PERIÓDICOS Y NOTICIAS

A fin de facilitar el tratamiento de la información, para la selección de periódicos se establecieron cuatro características: 1) edición en inglés, español, portugués o francés¹⁹; 2) cobertura, de forma habitual, del dominio política-economía. Así se asegura la coherencia semántica de términos, instancias y relaciones a identificar (Fernández-García et al., 2006); 3) tipo generalista, pues facilita la comprensión de las noticias²⁰; 4) Distribución en papel, pues aún tiene un peso importante y, además, el avance de Internet no es homogéneo.

Se dividió intencionalmente el mundo en nueve zonas (ver Figura 29), de acuerdo a sus competencias geopolíticas, históricas, culturales y socioeconómicas: Europa, Norteamérica, Latinoamérica, África subsahariana, antiguas repúblicas soviéticas, Asia, Oriente, Oceanía y Norte de África.

19 En Alemania no se halló ningún medio representativo no escrito en alemán.

20 En el caso de Australia, la ausencia de medios generalistas de referencia hizo que se optara por un medio de corte económico.

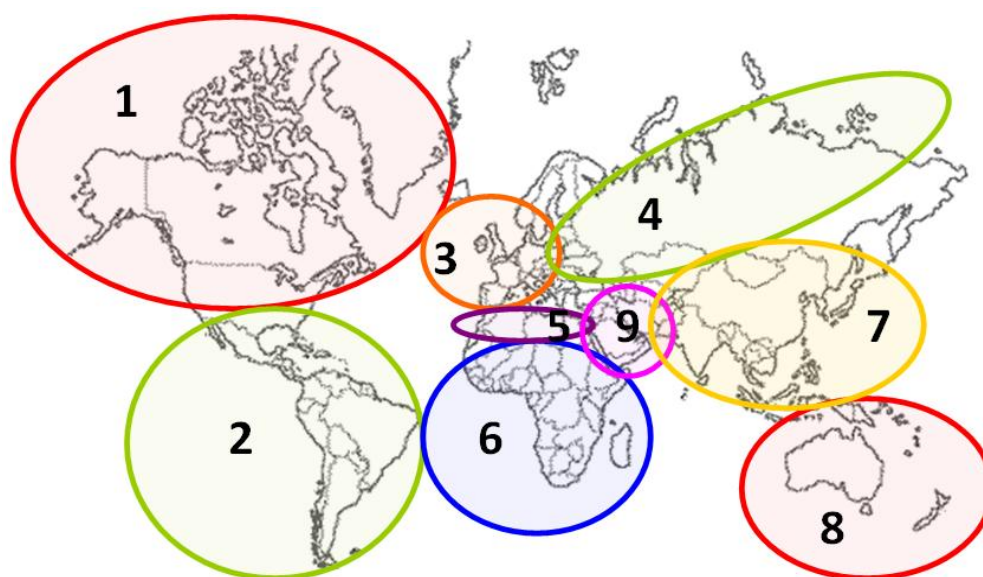


Figura 29. División del mundo por zonas geopolíticas

Esta división coincide, en términos generales, con las civilizaciones que Huntington (2001, pp. 35-42), identifica en el mundo y que define como entidades culturales, que engloban «aldeas, regiones, grupos étnicos, nacionalidades, grupos religiosos; todos tienen culturas distintas con diferentes grados de heterogeneidad cultural». Éstas pueden contener una o muchas unidades políticas, entre las que se incluyen: ciudades-estado, imperios, federaciones, confederaciones, estados-nación, estados multinacionales, y todas ellas pueden tener formas diversas de gobierno. Entre estas civilizaciones, Huntington distingue específicamente la china, la japonesa, la hindú, la islámica, la occidental, la latinoamericana y la africana. La islámica comprende, entre otras, las subcivilizaciones árabe, turca, persa y malaya; la occidental, Europa y Norteamérica, también hace referencia a sus antiguas colonias, Australia y Nueva Zelanda; Por lo que respecta a la africana señala que el «África subsahariana podría aglutinarse en una civilización peculiar, cuyo Estado central posiblemente sería Sudáfrica; el norte y la costa este pertenecen a la civilización islámica». Posteriormente, en cada área se seleccionaron los países con mayor Producto Interior Bruto (PIB), una medida macroeconómica que tiene en cuenta la producción de bienes y servicios de un país en un determinado momento.

Después se escogió el diario generalista más leído, de acuerdo a datos de accesos web y compra de edición impresa de 4IMN, que lleva a cabo un «ranking basado en la popularidad de sus páginas web, cuyo algoritmo incluye datos, imparciales e independientes»²¹. Parte de información de Google Page Rank, que asigna un peso numérico a cada elemento de un conjunto de documentos con hipervínculos; Majestic Seo Referring Domains, que mide la popularidad de un enlace a partir del número de links; Alexa Traffic Rank, cuya estrategia de medición de datos se basa en el uso que los usuarios

21 <http://www.4imn.com/> (último acceso: 23/05/2017)

4. RESULTADOS

hacen de su barra de herramientas y su extensión. 4IMN también muestra datos de ejemplares impresos vendidos. Para ello, utiliza como fuente la información proporcionada por organismos independientes que miden la *audiencia* de los periódicos.

Posteriormente, se estableció una doble puntuación para cada periódico, según datos de accesos a la Web y de difusión en papel. Teniendo en cuenta el total de periódicos de cada país, se asignó la mayor cifra, por ejemplo, 15, al periódico con mejores datos de acceso online; el siguiente valor, 14, al medio que ocupaba el segundo puesto, y así sucesivamente, hasta 1. Este mismo proceso se repitió en relación con los datos en papel. Después, se sumaron las dos puntuaciones, estableciéndose el ranking de los periódicos más populares en cada país. En caso de empate, se dio preferencia al medio con mayor puntuación en el número de accesos a la web del periódico. Si para un medio no había cifras de distribución, éste pasaba a ocupar las últimas posiciones de la tabla. La Tabla 2, con 21 periódicos, resume el proceso anterior.

Zona del mundo	País	Periódico	Código
Europa	Alemania	<i>Süddeutsche Zeitung</i>	101
	Francia	<i>Le Monde</i>	111
	Reino Unido	<i>The Daily Telegraph</i> <i>Financial Times</i> <i>The Economist</i>	121 122 123
Norteamérica	Estados Unidos	<i>The New York Times</i> <i>The Wall Street Journal</i>	201 202
Latinoamérica	Brasil	<i>O Globo</i>	301
	México	<i>El Univesal</i>	311
Antiguas repúblicas soviéticas	Rusia	<i>Pravda</i>	401
África Subsahariana	Nigeria	<i>Nigerian Tribune</i>	501
	Sudáfrica	<i>Independent Online</i>	511
Asia	China	<i>China Daily</i>	601
	India	<i>The Times of India</i>	611
	Japón	<i>The Asahi Shimbun</i>	621
Oriente	Arabia Saudí	<i>Arab News</i>	701
	Emiratos Árabes Unidos	<i>Gulf News</i>	711
	Israel	<i>Yedioth Aharonot</i>	721
	Turquía	<i>Today's Zaman</i>	731
Oceanía	Australia	<i>The Australian Financial Review</i>	801
Norte de África	Egipto	<i>The Daily News Egypt</i>	901

Tabla 2: Periódicos seleccionados

4. RESULTADOS

Por otro lado, por su trayectoria histórica y prestigio, los siguientes periódicos se incluyeron en la muestra, independientemente de los resultados de la selección muestral: a) *The New York Times* (USA). Fundado en 1851, de gran reputación, tomado frecuentemente como referencia, ha ganado más de noventa premios Pulitzer; b) *Financial Times* (Reino Unido). Nacido en 1.888, es considerado un diario de gran prestigio. Desde 1893 se empezó a imprimir en color salmón, lo que inició la tradición de usar este color como distintivo de la prensa económica; c) *The Economist* (Reino Unido), periódico económico internacional, de periodicidad semanal, se publica regularmente desde 1.843; d) *The Wall Street Journal* (USA), periódico financiero fundado en 1.889, ha recibido más de 30 Premios Pulitzer.

Diariamente y por cada periódico, de julio a diciembre de 2012 y 2013 se seleccionó y almacenó una noticia, conformando una muestra de 5.582 noticias. Las noticias fueron seleccionadas manualmente mediante acceso a la portada del medio. Es aquí donde suele aparecer la información considerada más relevante, destacando el titular principal en la parte superior de la página, con un tamaño de letra diferenciado y, a veces, por la fotografía (Fidalgo, 2004, p. 54). Si no era posible aplicar este criterio se seleccionaba la última noticia publicada en el momento de selección, o bien la que destacaran de forma generalizada el resto de medios internacionales. Si el periódico utilizaba titulares distintos en portada y en la noticia, se seleccionaba este último.

Los datos fueron almacenados mediante la extensión Clearly, de Evernote, en Google Chrome. Ésta permite almacenar el texto completo, cuando es posible, titular, fecha de publicación y url. Para facilitar la identificación del diario de procedencia, se utilizaron los códigos indicados en la Tabla 2. Una vez recopiladas las noticias, se descargaron en formato enex y se convirtieron a csv para facilitar su tratamiento.

4.1.3. DEFINICIÓN DE SUBMUESTRAS

Se definen dos muestras: una para el proceso de modelado de la ontología ONAPE y otra para su validación:

La submuestra₁ incluye artículos de cuatro meses, de julio a octubre de 2012, con un total de 1.837 noticias. Este conjunto de noticias, cuyo periodo fue seleccionado al azar, es utilizado para los procesos de indización humana y automática, permitiendo la modificación de ONA el modelado y modificación de ONAPE.

La submuestra₂ engloba diversos artículos, que fueron agrupados en cinco temas. Estos fueron seleccionadas aleatoriamente, a partir de las palabras clave que cada medio definió para describir las noticias en su código fuente (ver Tabla 3).

Palabra clave	Tema
Islas Senkaku (<i>Senkaku Islands</i>)	Conflicto territorial de las Islas Senkaku
Mandela	Homenajes a Nelson Mandela tras su muerte en 2013
Inmigración (<i>immigration</i>)	Restricciones a la entrada de inmigrantes en Reino Unido y Estados Unidos
Corrupción (<i>corruption</i>)	Caso de corrupción en Turquía
Energía (<i>energy</i>)	Mercado energético

Tabla 3: Temáticas de las noticias de la submuestra₂

Un resumen de cada tema y los titulares de las noticias seleccionadas como componentes de esta submuestra₂ aparece en el apartado 4.5.1. PREPROCESAMIENTO DE LA SUBMUESTRA₂ PARA LA EVALUACIÓN DE ONAPE: METADADOSHTML. Este conjunto de noticias es utilizado para la validación de la ontología ONAPE.

En el siguiente apartado se describe el proceso de modelado inicial de la ontología de dominio a partir de la Teoría de la Gramática de Casos. En esta etapa no se tiene en cuenta la muestra de periódicos y noticias aquí configurada.

4.2. MODELADO DE LA ONTOLOGÍA DE ACCIONES ONA BASADO EN LA GRAMÁTICA DE CASOS

Se detalla el procedimiento por el que los elementos de la Teoría de la Gramática de Casos de Fillmore (1967) y García Gutiérrez (2014) son adaptados para el modelado de ONA (ONTología de Acciones), orientada a la descripción de acciones, en general. Se entiende modelar como el proceso de organizar el conocimiento para su uso por parte de una comunidad (Allemang y Hendler, 2011, p. 28). Las características y componentes de este marco teórico son descritos en el apartado 2.3.3. INDIZACIÓN DE NOTICIAS. LA GRAMÁTICA DE CASOS de esta memoria.

4.2.1. ANÁLISIS DEL MARCO TEÓRICO Y ENTIDADES BÁSICAS

A continuación, se incide en el proceso de adaptación de los casos, que son clasificados en una o más entidades básicas. «Una entidad es una forma particular de categorización de tipo evento, lugar, organización, persona, producto y congéneres» (Rocha, Lessa y Monat, 2009). Una categoría, según estos autores, es «un recurso de clasificación que agrupa contenidos similares». En esta tesis acción y evento son términos equivalentes. La Tabla 4 muestra los casos y clasificación inicial en entidades básicas.

4. RESULTADOS

Casos	Acción	Sujeto	Objeto	Lugar	Fecha
01 Acción principal	X				
02 Objeto afectado por la acción			X		
03 Consecuencia, resultado de una acción	X				
04 Instrumento con que se realiza la acción			X		
05 Estado de quien ejecuta o realiza una acción			X		
06 Agentivo: Sujeto u Objeto que realiza una acción		X	X		
07 Dativo: ente animado afectado por la acción		X			
08 Acompañante: acompaña al agente en la acción		X			
09 Modo: Cómo se ejecuta la acción			X		
10 Lugar en que se produce la acción				X	
11 Fecha en que se produce la acción					X
12 Finalidad: propósito de una acción			X		
13 Causa: el por qué de una acción	X				

Tabla 4: Clasificación de casos en entidades básicas

Este planteamiento facilita, además, el establecimiento de relaciones entre dichos elementos. Por ejemplo, la relación entre un agente u objeto que realiza la acción y dicha acción puede ser “Agente ejecuta Acción”. «Personas, lugares y cosas son sustantivos, al fin y al cabo, con un rol esencial en el lenguaje que, con frecuencia, actúan como sujeto y/o objeto de la oración» (Ingersoll et al., 2013, p. 115).

La utilidad de la Gramática de Casos como marco teórico, en tanto que teoría basada en el análisis de roles de los términos en la oración, es que facilita la identificación y categorización de términos de documentos periodísticos en uno o más casos y su clasificación posterior en entidades básicas. Tanto los casos como las entidades básicas constituyen elementos potenciales de un vocabulario para la representación del contenido de esos documentos.

Las entidades básicas establecidas, que se definen a continuación, son comunes en otras ontologías, como FOAF (Brickley y Miller, 2014), Event (Raimond y Abdallah, 2007), Schema.org o Time (Hobbs y Pan, 2006). También se definen los casos que han sido clasificados en entidades básicas en la Tabla 4 y se incorporan ejemplos que muestran las posibles relaciones con y entre casos. Cada entidad básica y caso es una clase potencial de la ontología ONA. Los términos *Lugar* y *Fecha* han sido reemplazados por denominaciones que hacen referencia a conceptos más generales: *EspacioFisico* y *EspacioTemporal*, pues son denominaciones más amplias. El Diccionario de la Real Academia de la Lengua Española (en adelante, DLE) y/o García Gutiérrez (2014) son empleados como fuentes para las definiciones:

4. RESULTADOS

1. **Acción:** hecho que se produce en un espacio temporal y un espacio físico, en el que participa, al menos, un sujeto. El DLE define evento (sinónimo de acción en esta memoria) como «suceso importante y programado, de índole social, académica, artística o deportiva; situación descrita por un predicado, ya sea estática o dinámica» (RAE, 2014)²². Una de las dificultades de modelar acciones es que se fuerza la adopción de una perspectiva particular de lo que ocurre. Ejemplo: *Alberto Garzón presenta su candidatura* (acción) *para las elecciones*

Hay cierta falta de consenso sobre cómo establecer aserciones relativas a causalidad, propósito e influencia y su relación con las acciones. Algunas ontologías, como F, proponen una solución, que pasa por definir una restricción: sólo las acciones pueden mantener relaciones de causalidad, propósito y/o influencia con otras acciones (Shaw, Troncy y Hardman, 2009). Desde este punto de vista, una acción puede tener causas y consecuencias, que no son más que otras acciones relacionadas con la primera. A veces causa y consecuencia son acciones entre las que existe una relación inversa, pero no siempre.

- **Causa:** motivo que da lugar a la acción. Responde al por qué o al origen de ésta. Ejemplo: *Nathalie Arthaud ha sido investida Presidenta del Gobierno* (acción) *después de ganar las elecciones*.
 - **Consecuencia:** acción resultante de la ejecución de una acción. Ejemplo: *después de ganar las elecciones, Arthaud será investida en el Parlamento*.
2. **Sujeto:** «ser del cual se predica o anuncia algo» (RAE, 2014k)²³. Una persona es un sujeto, pero también lo son un grupo de personas (colectivos) o una organización. Los sujetos participan en la acción, de forma activa o pasiva, como:

- **Agente:** sujeto que ejecuta una acción. Ejemplo: *Mari Carmen investiga una fórmula para evitar disgustos*;
- **Receptor:** sujeto que recibe los efectos de una acción. Ejemplo: *Ángel contrató a Ginés para esta investigación*;
- **Acompañante:** sujeto que acompaña a quien ejecuta una acción. Ejemplo: *Miguel Ángel estaba con Lucía cuando redactó el contrato*.

En algunas ontologías de eventos, el agente, mediado por la propiedad *hasParticipant* conlleva un rol de actividad o causalidad.

3. **Objeto:** «todo lo que tiene entidad, ya sea corporal o espiritual, natural o artificial, real o abstracta» (RAE, 2014b)²⁴. Un objeto puede tener estas funciones:

22 Acepciones 3 y 4 del concepto evento.

23 Aceptación 6.

24 Aceptación 1 del concepto cosa.

4. RESULTADOS

- **Agente** que ejecuta una acción;
- **Objeto afectado:** objeto afectado por la acción ejecutada. Ejemplo: *Omar tramitó los documentos*;
- **Producto:** resultado de la ejecución de una acción. Ejemplo: *IBM fabrica buenos ordenadores*;
- **Instrumento:** de carácter tangible, «cosa o persona de que alguien se sirve para hacer algo o conseguir un fin» (RAE, 2014g)²⁵. Ejemplo: *Manuel aún escribe cartas con un lápiz*;
- **Estado:** intangible. Estado involuntario en el que puede encontrarse quien ejecuta o sufre la acción. Ejemplo: *cuando Pilar escribe se pone nerviosa*;
- **Modo:** de carácter intangible, hace referencia al procedimiento, técnica, manera, género con que se ejecuta una acción (García Gutiérrez, 2014). Ejemplo: *Fran escribe con mucha precisión*;
- **Finalidad:** objeto intangible. Deseos, objetivos, propósitos no alcanzados por los que ejecutar una acción; intención de quien ejecuta una acción. La diferencia entre finalidad y causa no sólo reside en la pregunta sino fundamentalmente en el carácter no consumado de la respuesta de la finalidad y el consumado de la causa, con datos y efectos comprobables en la consecuencia (García Gutiérrez, 2014). Ejemplo: *Mario escribía para acabar con la pobreza*.

4. **Espacio temporal:** espacio ocupado o que puede ser ocupado por un cuerpo cualquiera (definición basada en RAE (2014j), acepción 1 del concepto sitio). Es el espacio en que se produce una acción. Ejemplo: *en este país es habitual cambiar el sentido del voto*. Hay que tener en cuenta si estos espacios son reales o imaginarios, actuales o pasados ya que, si no existen en la actualidad, ¿son reales? Por otro lado, que un mismo lugar tenga un nombre diferente, de acuerdo con el momento histórico. También puede que sus fronteras hayan cambiado.

5. **Espacio físico:** «tiempo en que ocurre o se hace algo» (RAE, 2014e)²⁶, momento en que se produce la acción. Ejemplo: *las elecciones se celebrarán en 2017*. Para modelar el tiempo existen dos perspectivas: 1) definir rangos de tiempo mediante literales (propiedades dato); 2) clases que representen estos intervalos temporales.

Entre estas clases existe una serie de propiedades tipo objeto (relaciones) que las vinculan. De esta forma, se pueden establecer el dominio y el rango de cada propiedad (ver la Figura 30). Por lo que respecta a su formalización, ya se emplea la notación *CamelBack*, y en la medida de lo posible, se siguen las recomendaciones de Aguado-de-Cea et al. (2015):

25 Acepción 2.

26 Acepción 2 del concepto fecha.

4. RESULTADOS

1. *acompañaA*: relaciona acompañante y agente de la acción;
2. *afectaA*: relaciona acción y receptor de su ejecución;
3. *ejecuta*: relaciona el agente de una acción y ésta;
4. *esAcompañadoPor*: relaciona el agente que ejecuta una acción y su acompañante;
5. *esAfectadoPor*: relaciona receptor de una acción y ésta;
6. *esCausaDe*: relaciona causa y acción;
7. *esConsecuenciaDe*: relaciona acción y su consecuencia;
8. *esEjecutadaPor*: relaciona acción y agente que la ejecuta;
9. *esEspacioFisicoDe*: relaciona un espacio físico y la acción;
10. *esEspacioTemporalDe*: relaciona un espacio temporal y una acción que se produce en dicho espacio;
11. *esSituacionDe*: relaciona el estado involuntario en el que puede encontrarse quien ejecuta o recibe los efectos de una acción con dicho sujeto;
12. *esFinalidadDe*: relaciona finalidad de un agente y la acción que lleva a cabo;
13. *esModoDe*: relaciona el modo en que se ejecuta una acción y dicha acción;
14. *esProductoDe*: relaciona resultado de una acción y ésta;
15. *instrumentoUsadoEn*: relaciona un instrumento y la acción en que es utilizado;
16. *tieneCausa*: relaciona una acción y la causa por la que se ejecuta;
17. *tieneConsecuencia*: relaciona acción y consecuencia;
18. *tieneEspacioFisico*: relaciona una acción y el espacio físico en el que se produce;
19. *tieneEspacioTemporal*: relaciona acción y el espacio temporal en que se produce;
20. *tieneSituacion*: relaciona un sujeto y el estado involuntario en el que se encuentra cuando ejecuta o recibe los efectos de una acción;
21. *tieneFinalidad*: relaciona una acción y el propósito para la que se ejecuta;
22. *tieneInstrumento*: relaciona una acción y el instrumento usado en ésta;
23. *tieneModo*: relaciona una acción y el modo en que se ejecuta;
24. *tieneProducto*: relaciona acción y su producto o consecuencia.

El dominio y el rango de las relaciones son perceptibles con mayor facilidad en la Figura 30, donde se representan gráficamente las clases y relaciones definidas anteriormente. Cada clase obtenida a partir de una entidad básica emplea un color diferente: naranja (sujeto), gris (acción), verde (objeto), azul (espacio temporal) y rosa (espacio físico). Los casos aparecen en amarillo. Por otro lado, para facilitar la lectura de la figura anterior, la clase objeto aparece dos veces. Para las relaciones existen relaciones inversas que, para facilitar la visualización del modelo, no se han añadido.

4.2.2. MODELADO INICIAL DE ONA

La Gramática de Casos se define a nivel de lenguaje natural. Sin embargo, las ontologías lo hacen a nivel formal. Se analiza y modifica el modelo anterior con vistas a su formalización en OWL. Así, se plantea una serie de cuestiones para facilitar la correcta identificación de relaciones entre clases y la separación entre instancias y su participación en una acción específica.

Por un lado, se han invertido las denominaciones de *Sujeto* y *Agente*, con el propósito de facilitar la interpretación de las acciones y su relación con el resto de elementos. En la Figura 31 se muestran tres clases (*Accion*, *Sujeto* y *Objeto*) y dos relaciones (*esSujetoDe*, *esObjetoDe*). Se crea así una estructura similar a las tripletas RDF (sujeto-predicado-objeto).



Figura 31: Relaciones básicas del modelo

Para evitar problemas de homonimia en relación con el término *Estado* como situación y *Estado* como «conjunto de instituciones que poseen la autoridad y potestad para establecer las normas que regulan una sociedad, teniendo soberanía interna y externa sobre un País», también se ha reemplazado *Estado* por *Situacion*.

Por otro lado, se han modificado las relaciones entre *Instrumento*, *Agente* y *Accion* con respecto a cómo aparecen en la Figura 30. Tal como son definidas por la Gramática de Casos no es posible identificar correctamente la función de cada elemento en una *Accion*, por lo que se plantean tres soluciones posibles:

- a) Modificación del modelo estableciendo la participación del *Instrumento* directamente con la acción (ver Figura 32) es insuficiente. No se está relacionando *Agente* e *Instrumento*. En caso de que participaran dos *Agentes*, ¿quién usa el *Instrumento*?

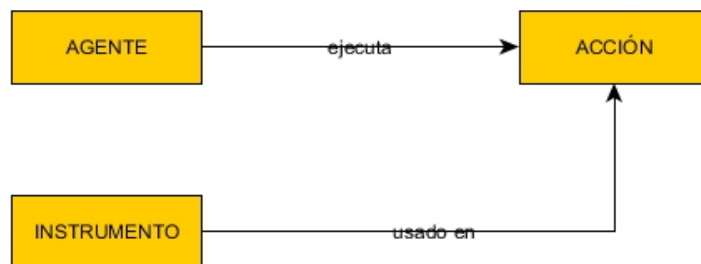


Figura 32: a) Relaciones Agente-Accion e Instrumento-Accion

4. RESULTADOS

- b) Modificación del modelo manteniendo la relación *Agente-Instrumento* y añadiendo la propiedad tipo objeto *Instrumento-Accion* (ver Figura 33) no mantendría la independencia del *Agente* respecto del *Instrumento* con respecto a su participación en otras acciones. Si un *Agente* ha sido relacionado con dos *Instrumentos*, en dos instancias de *Accion*, ¿con qué *Instrumento* ha participado en cada una?

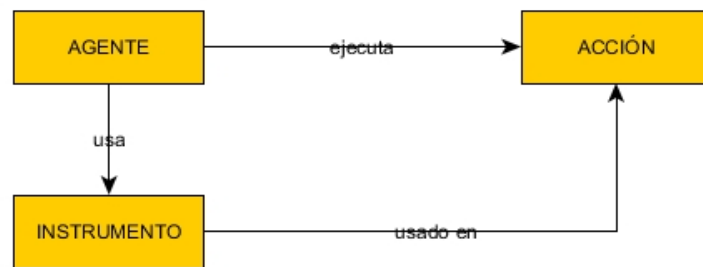


Figura 33: b) Relaciones *Agente-Accion*, *Agente-Instrumento* e *Instrumento-Accion*

- c) Incorporación de un nodo anónimo o clase auxiliar que es el que mantiene las relaciones entre los tres elementos indicados conseguiría expresar la participación de un *Agente* en una *Accion* con un *Instrumento* específico, por un lado y asegurar la independencia de la instancia del *Agente* para cualquier otro tipo de *Accion*, ya que no estaría vinculada directamente con ningún *Instrumento*. De esta forma, una clase auxiliar, que en la Figura 34 se ha denominado *AgenteConInstrumento* a modo de ejemplo, es la que *ejecutaría* la *Accion*.

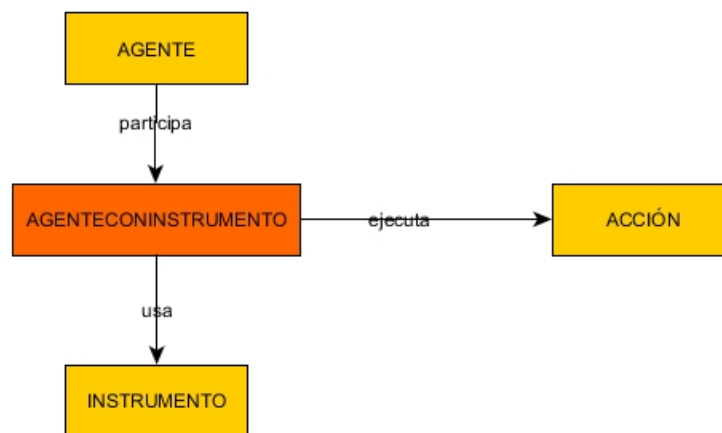


Figura 34: c) Relaciones de una clase auxiliar con *Agente*, *Instrumento* y *Accion*

Esta solución también se aplica a otras instancias de clases del modelo que, como *Instrumento*, se ven afectadas por las cuestiones indicadas: *Modo*, *Finalidad* y *Estado*. De esta forma una clase anónima, que se ha denominado *Sujeto*, es quien ejecuta la *Accion* y, a la vez, relaciona *Agente* con el *Estado* en que se encuentra, el *Instrumento* que emplea, el *Modo* en que realiza la *Acción* y con qué *Finalidad* (ver Figura 35).

4. RESULTADOS

De forma similar, *Objeto* es un elemento auxiliar que relaciona el *Receptor* de una *Accion* y el *Estado* en que se encuentra cuando recibe los efectos de la misma.

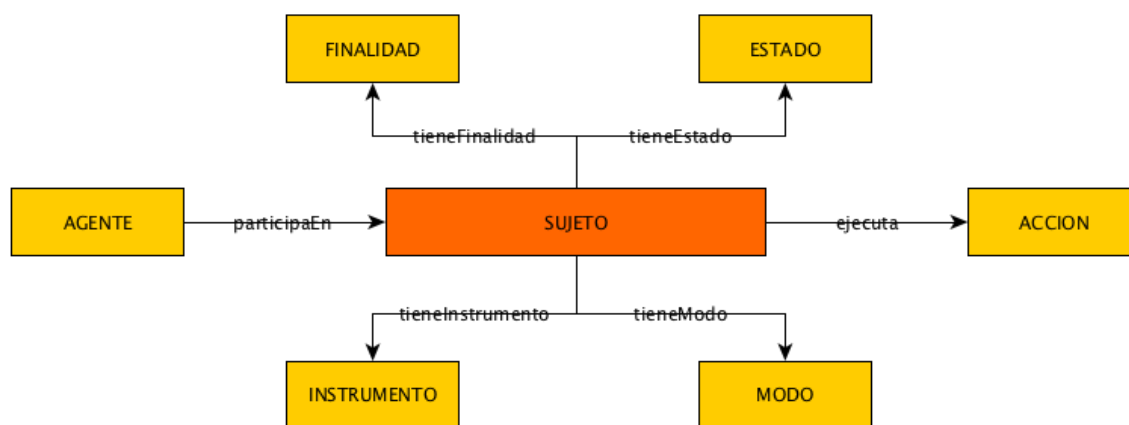


Figura 35: Relación entre *AgenteAccionEjecutada* y resto de clases relacionadas

Por último, se han suprimido las clases *Finalidad*, *Estado*, *Instrumento*, *Acompañante*, *Causa* y *Producto*. Este tipo de información es representado únicamente mediante propiedades tipo objeto. Con ello, se simplifica el modelo, facilitando su interpretación.

Las equivalencias entre casos y elementos de ONA se indican en la siguiente Tabla.

Casos	Elementos del modelo
Acción	<i>Accion</i>
Objetivo	<i>Objeto (Cosa)</i>
Factitivo	<i>esResultadoDe, tieneResultado</i>
Instrumental	<i>esInstrumentoDe, tieneInstrumento</i>
Estado	<i>esSituaciónDe, tieneSituacion</i>
Agentivo	<i>Sujeto (Agente, Cosa)</i>
Dativo	<i>Objeto (Agente)</i>
Asociativo	<i>esAcompañanteDe, tieneAcompañante</i>
Modo	<i>esModoDe, tieneModo</i>
Locativo	<i>EspacioTemporal esEspacioTemporalDe, tieneEspacioTemporal</i>
Fecha	<i>EspacioFisico esEspacioFisicoDe, tieneEspacioFisico</i>
Causativo	<i>esCausaDe, tieneCausa</i>
Finalidad	<i>esFinalidadDe, tieneFinalidad</i>

Tabla 5: Equivalencias entre Casos y elementos de ONA

4. RESULTADOS

Como ya se apuntaba antes, se sigue la notación *CamelBack* para las denominaciones de clases y relaciones. También las recomendaciones de Aguado-de-Cea et al. (2015), tanto para las denominaciones como para la descripción de ONA en la Tabla 6 y Tabla 7. Los términos, inicialmente en español, fueron traducidos al inglés, utilizando la API de Google Translate²⁷. Posteriormente fueron revisados usando Word Reference²⁸ y Linguee²⁹.

Clase	Descripción
Accion (Action)	Entidad básica que expresa el hecho que se produce en un <i>EspacioTemporal</i> y un <i>EspacioFisico</i> (conocidos o no), en el que puede participar, como mínimo, un <i>Sujeto</i> Es dominio de: <i>esCausaDe</i> , <i>esConsecuenciaDe</i> , <i>tieneCausa</i> , <i>tieneConsecuencia</i> , <i>tieneEspacioFisico</i> , <i>tieneEspacioTemporal</i> , <i>tieneObjeto</i> , <i>tieneResultado</i> , <i>tieneSujeto</i> Es rango: <i>esCausaDe</i> , <i>esConsecuenciaDe</i> , <i>esEspacioFisicoDe</i> , <i>esEspacioTemporalDe</i> , <i>esObjetoDe</i> , <i>esResultadoDe</i> , <i>esSujetoDe</i> , <i>tieneCausa</i> , <i>tieneConsecuencia</i>
Agente (Agent)	Ser que puede participar en una <i>Accion</i> , llevándola a cabo, acompañando a quien la ejecuta o recibiendo sus efectos Es dominio de: <i>esAcompañanteDe</i> , <i>esEjecutanteDe</i> , <i>esReceptorDe</i> , <i>tieneFinalidad</i> Es rango de las relaciones: <i>esFinalidadDe</i> , <i>tieneAcompañante</i> , <i>tieneEjecutante</i> , <i>tieneReceptor</i>
Cosa (Stuff)	Todo lo que tiene entidad, tangible o intangible Es dominio: <i>esAcompañanteDe</i> , <i>esEjecutanteDe</i> , <i>esFinalidadDe</i> , <i>esInstrumentoDe</i> , <i>esModoDe</i> , <i>esReceptorDe</i> , <i>esResultadoDe</i> , <i>esSituacionDe</i> Es rango: <i>tieneAcompañante</i> , <i>tieneEjecutante</i> , <i>tieneFinalidad</i> , <i>tieneInstrumento</i> , <i>tieneModo</i> , <i>tieneReceptor</i> , <i>tieneResultado</i> , <i>tieneSituacion</i>
EspacioFisico (PhysicalSpace)	Espacio físico o lugar en el que tiene lugar una <i>Accion</i> Es dominio: <i>esEspacioFisicoDe</i> , <i>tieneReceptor</i> Es rango: <i>tieneEspacioFisico</i> , <i>esReceptorDe</i>
EspacioTemporal (TemporalSpace)	<i>EspacioTemporal</i> (fecha, periodo, tiempo) en el que tiene lugar una <i>Accion</i> Es dominio de las relaciones: <i>esEspacioTemporalDe</i> , <i>tieneReceptor</i> Es rango de las relaciones: <i>tieneEspacioTemporal</i> , <i>esReceptorDe</i>
Objeto (Object)	Clase auxiliar que vincula un <i>Agente</i> o <i>Cosa</i> y la <i>Accion</i> de la que recibe los efectos de su ejecución. Esta clase “artificial” es utilizada para vincular todos los elementos relacionados con el <i>Agente</i> o <i>Cosa</i> en la recepción de los efectos de la <i>Accion</i> . Es dominio: <i>esObjetoDe</i> , <i>tieneReceptor</i> , <i>tieneSituacion</i> Es rango: <i>tieneObjeto</i> , <i>esReceptorDe</i> , <i>esSituacionDe</i>
Sujeto (Subject)	Clase auxiliar que vincula un <i>Agente</i> o <i>Cosa</i> y la <i>Accion</i> que ejecuta. Es utilizada para vincular todos los elementos relacionados con un <i>Agente</i> en la ejecución de una <i>Accion</i> . Es dominio: <i>tieneSituacion</i> , <i>esSujetoDe</i> , <i>tieneAcompañante</i> , <i>tieneEjecutante</i> , <i>tieneFinalidadDe</i> , <i>tieneInstrumento</i> , <i>tieneModo</i> Es rango: <i>esAcompañanteDe</i> , <i>esEjecutanteDe</i> , <i>esSituacionDe</i> , <i>esFinalidadDe</i> , <i>esInstrumentoDe</i> , <i>esModoDe</i> , <i>tieneSujeto</i>

Tabla 6: Clases de ONA

27 <https://cloud.google.com/translate/docs> (último acceso: 30/04/2017)

28 <http://www.wordreference.com> (último acceso: 09/07/2017)

29 <http://www.linguee.es> (último acceso: 30/04/2017)

4. RESULTADOS

Las relaciones entre clases se muestran en la Tabla 7. Se indica relación, su descripción, su dominio y rango.

Propiedad	Descripción	Dominio	Rango
esAcompañanteDe (isCompanionOf)	Relaciona <i>Agente</i> o <i>Cosa</i> y <i>Sujeto</i> . Inversa: <i>tieneAcompañante</i>	<i>Agente</i> <i>Cosa</i>	<i>Sujeto</i>
esCausaDe (isCauseOf)	Relaciona una <i>Accion</i> que <i>esCausaDe</i> otra y dicha <i>Accion</i> . Inversa: <i>tieneCausa</i>	<i>Accion</i>	<i>Accion</i>
esConsecuenciaDe (isConsequenceOf)	Relaciona la <i>Accion</i> consecuencia de una <i>Accion</i> y ésta. Inversa: <i>tieneConsecuencia</i>	<i>Accion</i>	<i>Accion</i>
esEjecutanteDe (isExecutorOf)	Relaciona el <i>Agente</i> o <i>Cosa</i> que ejecuta una <i>Accion</i> y el <i>Sujeto</i> a través del que la ejecuta. Inversa: <i>tieneEjecutante</i>	<i>Agente</i> <i>Cosa</i>	<i>Sujeto</i>
esEspacioFisicoDe (isPhysicalSpaceOf)	Relaciona el <i>EspacioFisico</i> en que se produce una <i>Accion</i> y dicha <i>Accion</i> . Inversa: <i>tieneEspacioFisico</i>	<i>EspacioFisico</i>	<i>Accion</i>
esEspacioTemporalDe (isTemporalSpaceOf)	Relaciona el <i>EspacioTemporal</i> en que se produce una <i>Accion</i> y dicha <i>Accion</i> . Inversa: <i>tieneEspacioTemporal</i>	<i>EspacioTemporal</i>	<i>Accion</i>
esFinalidadDe (isPurposeOf)	Relaciona la intención de un <i>Agente</i> en la ejecución de una <i>Accion</i> por medio de un <i>Sujeto</i> y ese <i>Sujeto</i> . Inversa: <i>tieneFinalidad</i>	<i>Cosa</i>	<i>Sujeto</i>
esInstrumentoDe (isInstrumentOf)	Relaciona un <i>Sujeto</i> y el <i>Instrumento</i> que un <i>Agente</i> usa para ejecutar una <i>Accion</i> . Inversa: <i>tieneInstrumento</i>	<i>Cosa</i>	<i>Sujeto</i>
esModoDe (isModeOf)	Relaciona una <i>Cosa</i> utilizada como <i>Instrumento</i> para la ejecución de una <i>Accion</i> y el <i>Sujeto</i> a través del que es utilizada. Inversa: <i>tieneModo</i>	<i>Cosa</i>	<i>Sujeto</i>
esObjetoDe (isObjectOf)	Relaciona el <i>Objeto</i> a través del que un <i>Agente</i> o <i>Cosa</i> recibe los efectos de una <i>Accion</i> y dicha <i>Accion</i> . Inversa: <i>tieneObjeto</i>	<i>Objeto</i>	<i>Accion</i>
esReceptorDe (isReceiverOf)	Relaciona el elemento, habitualmente un <i>Agente</i> o <i>Cosa</i> , que recibe los efectos de una <i>Accion</i> y <i>Objeto</i> , clase auxiliar a través del que los recibe. Inversa: <i>tieneReceptor</i>	<i>Objeto</i>	<i>Agente</i> <i>Cosa</i>
esResultadoDe (isResultOf)	Relaciona la <i>Cosa</i> producto de la ejecución de una <i>Accion</i> y ésta. Inversa: <i>tieneResultado</i>	<i>Cosa</i>	<i>Accion</i>
esSituacionDe (isFeelingOf)	Relaciona el estado en que se encuentra un <i>Agente</i> o <i>Cosa</i> y éste, a través de un <i>Sujeto</i> u <i>Objeto</i> , en la ejecución o recepción de una <i>Accion</i> . Inversa: <i>tieneSituacion</i>	<i>Cosa</i>	<i>Sujeto</i> <i>Objeto</i>
esSujetoDe (isSubjectOf)	Relaciona <i>Sujeto</i> y la <i>Accion</i> que ejecuta. Inversa: <i>tieneSujeto</i>	<i>Sujeto</i>	<i>Accion</i>
tieneAcompañante (hasCompanion)	Relaciona un <i>Sujeto</i> y el <i>Agente</i> o <i>Cosa</i> que acompaña a otro <i>Agente</i> o <i>Cosa</i> en la ejecución de una <i>Accion</i> . Inversa: <i>esAcompañanteDe</i>	<i>Sujeto</i>	<i>Agente</i> <i>Cosa</i>

4. RESULTADOS

Propiedad	Descripción	Dominio	Rango
tieneCausa (hasCause)	Propiedad que relaciona una <i>Accion</i> y la <i>Causa</i> por la que ocurre. Inversa: <i>esCausaDe</i>	<i>Accion</i>	<i>Accion</i>
tieneConsecuencia (hasConsequence)	Propiedad que relaciona una <i>Accion</i> y la <i>Consecuencia</i> que produce. Inversa: <i>esConsecuenciaDe</i>	<i>Accion</i>	<i>Accion</i>
tieneEjecutante (hasExecutor)	Relaciona un <i>Sujeto</i> y el <i>Agente</i> o <i>Cosa</i> que ejecuta una <i>Accion</i> a través del mismo. Inversa: <i>esEjecutanteDe</i>	<i>Sujeto</i>	<i>Agente</i> <i>Cosa</i>
tieneEspacioFisico (hasPhysicalSpace)	Relaciona una <i>Accion</i> y el <i>EspacioFisico</i> en que se produce. Inversa: <i>esEspacioFisicoDe</i>	<i>Accion</i>	<i>EspacioFisico</i>
tieneEspacioTemporal (hasTemporalSpace)	Relaciona una <i>Accion</i> y el <i>EspacioTemporal</i> en que se produce. Inversa: <i>esEspacioTemporalDe</i>	<i>Accion</i>	<i>EspacioTemporal</i>
tieneFinalidad (hasPurpose)	Relaciona un <i>Sujeto</i> y la finalidad o intención de un <i>Agente</i> (a través de ese <i>Sujeto</i>) por la que ejecuta una <i>Accion</i> . Es entendido como objetivo, propósito o intención. Inversa: <i>esFinalidadDe</i>	<i>Sujeto</i>	<i>Cosa</i>
tieneInstrumento (hasInstrument)	Relaciona un <i>Agente</i> que ejecuta una acción, a través de <i>Sujeto</i> y el <i>Instrumento</i> que utiliza para llevarla a cabo. Inversa: <i>esInstrumentoDe</i>	<i>Sujeto</i>	<i>Cosa</i>
tieneModo (hasMode)	Relaciona el <i>Modo</i> en que un <i>Agente</i> , a través de la clase <i>Sujeto</i> , ejecuta una <i>Accion</i> , y ese <i>Sujeto</i> . Inversa: <i>esModoDe</i>	<i>Sujeto</i>	<i>Cosa</i>
tieneObjeto (hasObject)	Relaciona <i>Accion</i> y el <i>Objeto</i> por medio del cual un <i>Agente</i> o <i>Cosa</i> recibe sus efectos. Inversa: <i>esObjetoDe</i>	<i>Accion</i>	<i>Objeto</i>
tieneReceptor (hasReceiver)	Relaciona <i>Objeto</i> , como clase auxiliar, con el <i>Agente</i> o <i>Cosa</i> que recibe el efecto de una <i>Accion</i> . Inversa: <i>esReceptorDe</i>	<i>Objeto</i>	<i>Agente</i> <i>Cosa</i>
tieneResultado (hasResult)	Relaciona la <i>Cosa</i> resultante de la ejecución de una <i>Accion</i> y dicha <i>Accion</i> . Inversa: <i>esResultadoDe</i>	<i>Accion</i>	<i>Cosa</i>
tieneSituacion (hasFeeling)	Relaciona el <i>Sujeto</i> u <i>Objeto</i> y la situacion en que se encuentra en la ejecución o recepción de una <i>Accion</i> . Inversa: <i>esSituacionDe</i>	<i>Sujeto</i> <i>Objeto</i>	<i>Cosa</i>
tieneSujeto (hasSubject)	Relaciona una <i>Accion</i> y el <i>Sujeto</i> a través del que un <i>Agente</i> o <i>Cosa</i> la ejecuta. Inversa: <i>esSujetoDe</i>	<i>Accion</i>	<i>Sujeto</i>

Tabla 7: Relaciones iniciales de ONA

A las anteriores hay que añadir dos relaciones, de tipo jerárquico: las de clase-subclase y las enumerativas, entre clases e instancias. Estas propiedades no se definen explícitamente en Protégé, ya que este software las incorpora gráficamente.

Otros aspectos del modelo ONA son:

1. Todos los elementos son repetibles y opcionales, a excepción de *Accion*, que es elemento central y, por tanto, obligatorio
2. Una clase *Agente* o *Cosa* no realiza una *Accion* directamente, sino a través de una clase auxiliar *Sujeto*. Así, se facilita la relación entre quién o qué ejecuta una acción y el resto de elementos que afectan a la forma en que ejecuta una acción, por ejemplo, su finalidad, acompañante, instrumento, etc.
3. Una clase *Agente* o *Cosa* no recibe los efectos de una *Accion* directamente, sino a través de una clase auxiliar *Objeto*. De esta forma, se facilita la relación entre quién o qué recibe los efectos de la ejecución de una acción y el resto de cuestiones que inciden en cómo la recibe, por ejemplo, su situación o acompañante, etc.

La Figura 36 y la Figura 37 representan el modelo ONA. Las clases básicas aparecen en amarillo y las auxiliares en naranja. La Figura 37 muestra las relaciones inversas con respecto de las indicadas en la Figura 36. Este modelo permite responder a las 12W de García Gutiérrez (2014), que cubre los casos de la Gramática de Casos de Fillmore (1967). Con ello se atienden, por tanto, las funciones conceptuales básicas (casos) que pueden aparecer en una noticia y facilita la representación del contenido de documentos cuyo principal componente son acciones.

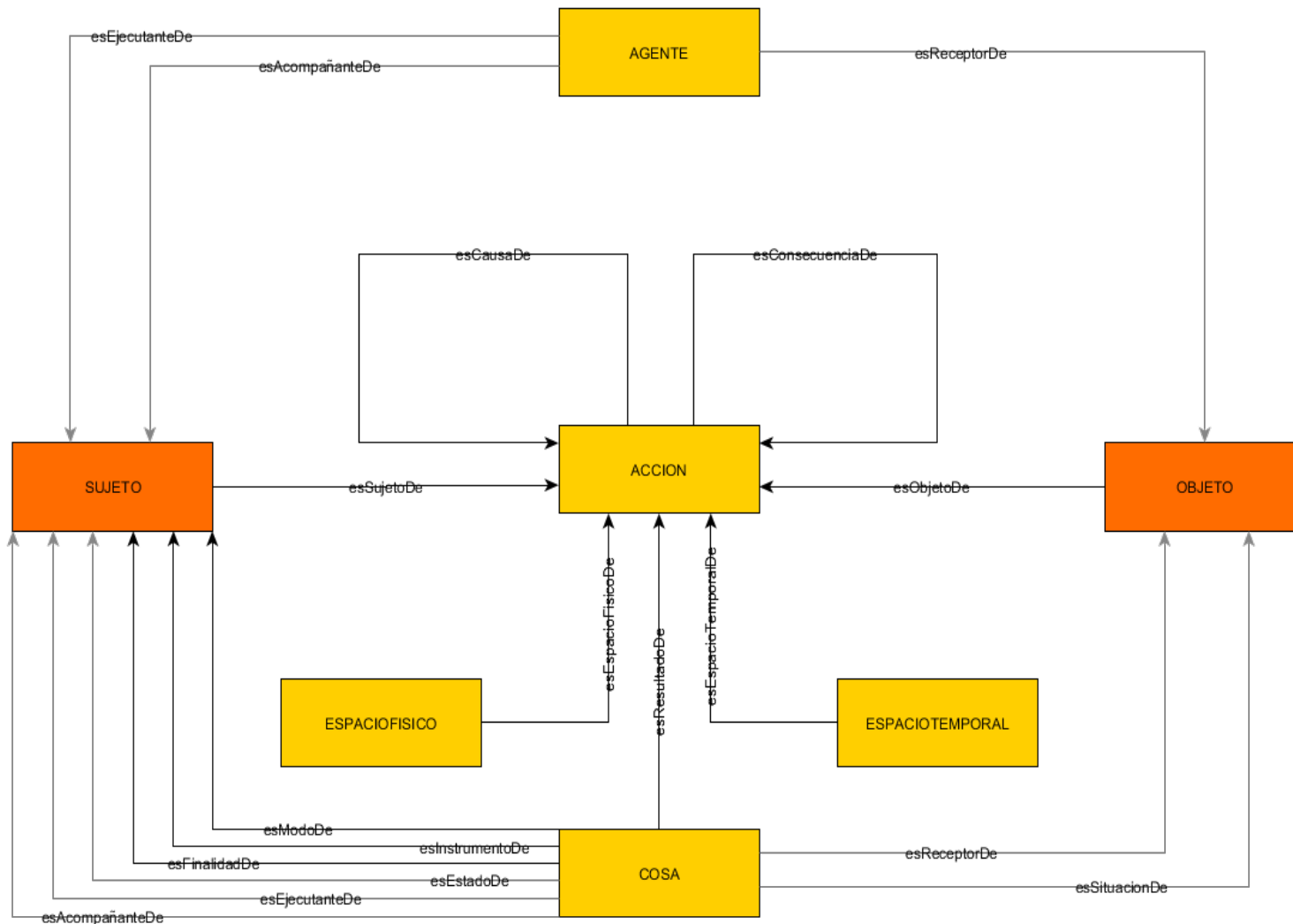


Figura 36: Modelo ONA inicial

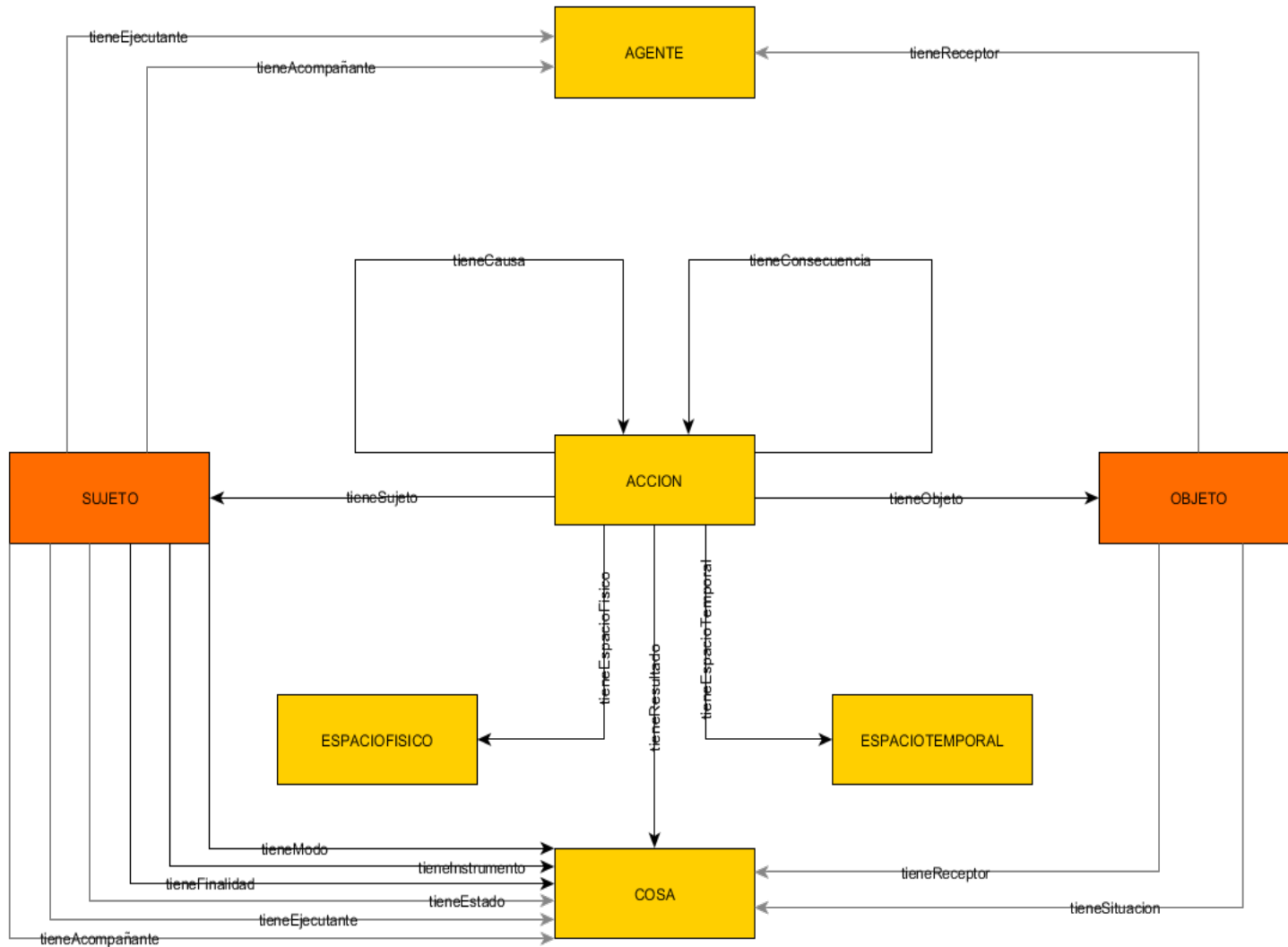


Figura 37: Modelo ONA inicial - Relaciones inversas

ONA es formalizada en OWL en el software Protégé, descrito en 2.2.4. HERRAMIENTAS PARA ONTOLOGÍAS: PROTÉGÉ.

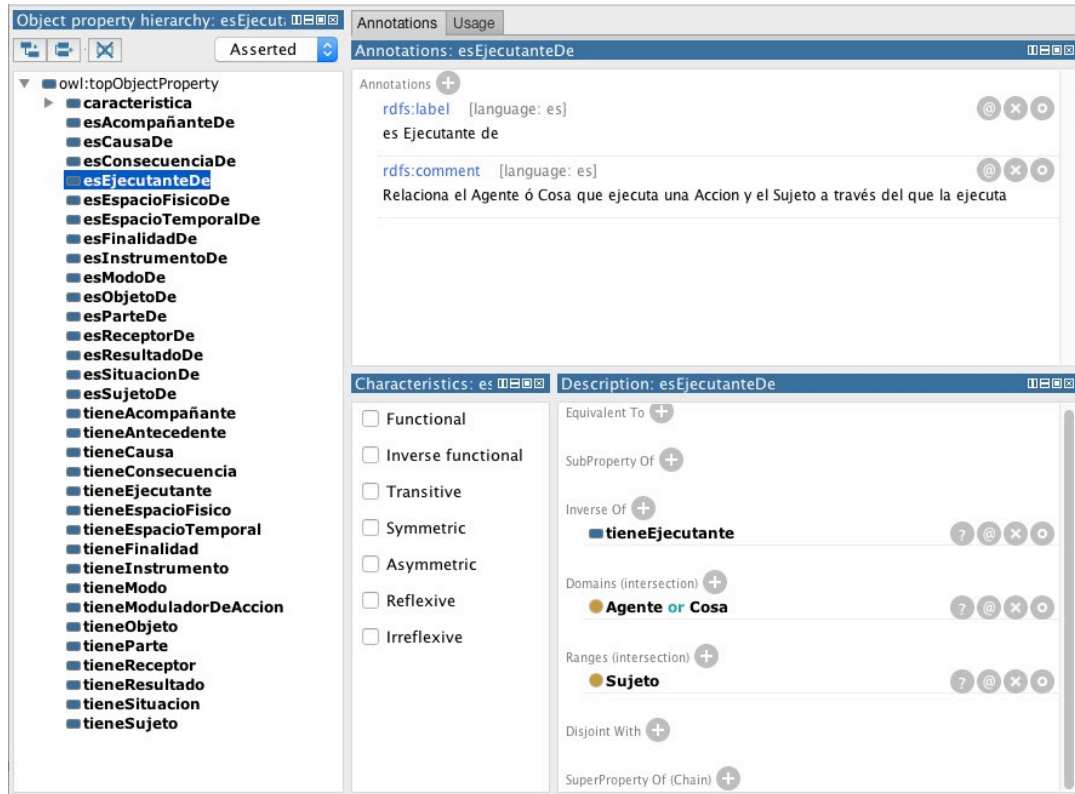


Figura 38: Relación *esEjecutanteDe* y las anotaciones "label" y "comment" en Protégé

Las características de los elementos de ONA se recogen en propiedades de anotación (*annotation*) (véase la Figura 38): para las definiciones, se emplea la propiedad *comment*; para las denominaciones, la propiedad *label*; para fuentes de las definiciones, *isDefinedBy*.

El anexo B. ESPECIFICACIÓN DE ONA describe el modelo ONA y recoge su representación gráfica final, incluyendo los elementos incorporados en la siguiente etapa.

4.2.3. DATOS ESTADÍSTICOS DEL PROCESO

Los datos estadísticos de este proceso son los que se indican a continuación: 20 clases; 32 propiedades tipo objeto (relaciones), con sus respectivos dominios y rangos y 23 individuos. Además, se han definido 16 propiedades inversas; 196 propiedades de anotación, de las cuales 75 son *comments*, 40 de definición (*isDefinedBy*) y 75 son *labels*.

ONA es el punto de partida para el modelado de acciones de política-economía mediante, ONAPE (ONtología de Acciones de Política y Economía). Para la obtención de términos susceptibles de ser incorporados a este modelo, se aplica indización humana y automática sobre la submuestra₁ ya descrita. Además, ONA es completada con nuevos elementos.

4.3. DESCUBRIMIENTO DE CONOCIMIENTO POR INDIZACIÓN. MODELADO DE LAS ONTOLOGÍAS

Obtenido el modelo inicial de la ontología a partir de la Gramática de Casos de Fillmore (1967) y García Gutiérrez (2014), ONA, se lleva a cabo la indización humana y automática de la submuestra₁. Se obtiene así un conjunto de términos con los que modelar ONAPE, como ontología del dominio política-economía, y con los que modificar ONA, en los casos en que procede. Aunque la indización automática permite extraer términos de distintos *corpora*, aún tiene problemas para juzgar si una unidad léxica es relevante en determinado contexto. Así, si bien la indización humana es un proceso lento y costoso, puede resolver parte de esta problemática, pues permite detectar matices y elementos contextuales (procesamiento pragmático) que la máquina aún no puede hallar. Además, facilita la identificación de términos que, estando presentes de alguna forma en el documento, no aparecen explícitamente. Se trata, pues, de aprovechar las ventajas y características de cada una (Anderson y Pérez-Carballo, 2001a).

4.3.1. ATRIBUCIÓN DE TÉRMINOS POR INDIZACIÓN HUMANA

Este procedimiento consiste en el acceso a cada elemento de la submuestra₁ y, siguiendo a Anderson y Pérez-Carballo (2001a) y Lancaster (2004, p. 9), analizar el documento para determinar su asunto y traducirlo en un conjunto de términos que representen su contenido. En esta traducción se pueden emplear distintos términos o materias de documentos (Taylor, 1999, p. 137), algunos extraídos, otros atribuidos. Esta atribución se basa en criterios como la experiencia previa, la aparición en el titular o en noticias sobre la misma cuestión.

En total se indizan 1.837 items (submuestra₁), asignando a éstos diversos términos, representantes de determinados conceptos en el contexto político-económico. De esta forma, se obtiene un primer listado de términos susceptibles de ser incorporados a ONAPE. Entre estos términos aparecen entidades nombradas (instancias en la ontología) y se identifican algunos temas relevantes. Estos últimos son frecuentes en el ámbito periodístico, donde «dependiendo de la emergencia o de la búsqueda, un tema hasta ahora inexistente puede ser destacado (...), siendo incorporados a su estructura de navegación, lo que es característico de los *breaking news*» (Rocha et al., 2009). Inicialmente se obtuvieron 663 términos diferentes. La Tabla 8 es una muestra de los términos resultantes.

País	Banco	Minoría étnica	Conflicto	Energía
Estados Unidos	HSBC	Rohinyas	Guerra de Yom Kippur	Energía nuclear
Libia	Barclays	Gitanos	Mar de la China Meridional	Energía renovable
Sudán	Banksia	Kurdos	Cachemira	Energía solar

Tabla 8: Palabras clave de descripción de noticias obtenidas por indización humana

La primera fila de la Tabla anterior muestra el concepto o clase, el resto de elementos son instancias, a partir de las cuales se han identificado las clases indicadas. La última columna, sombreada, muestra subclases de Energía.

4.3.2. EXTRACCIÓN DE TÉRMINOS POR INDIZACIÓN AUTOMÁTICA

Para la indización automática se utiliza e-Termos (*Termos Eletrônicos*), que funciona por etapas (ver 2.3.2. EXTRACCIÓN DE TÉRMINOS CON N-GRAMAS: e-TERMOS. ANÁLISIS MORFOLÓGICO: GRAMPAL). En la segunda etapa del software, se cargan los titulares de la submuestra₁, identificando género (informativo) y tipo de texto (noticia). Después, se compilan en un corpus de trabajo, de cara a la siguiente etapa. Los datos estadísticos iniciales se indican en la Tabla 9.

TIPO DE DATO ESTADÍSTICO	CIFRAS TOTALES
Palabras halladas en los titulares	21396
Palabras diferentes	5580
Siglas, acrónimos y nombres propios posibles encontrados	847

Tabla 9: Datos estadísticos iniciales obtenidos en e-Termos

En la tercera etapa (ver Figura 39), se extraen listas de términos, con combinaciones de 1 a 7 gramas (1 grama = 1 término). Se desecharon aquellos elementos sin relevancia semántica (palabras vacías), empleando una lista de palabras vacías (artículos, pronombres, preposiciones, etc.). De esta forma, además de reducir la cantidad de falsos positivos, se agiliza el proceso de indización (Anderson y Pérez-Carballo, 2001b). Aunque, en principio, la técnica n-gramas puede aplicarse independientemente del idioma (Martínez-Fernández et al., 2004), para utilizar la lista de palabras vacías es necesario trabajar en una lengua. Por este motivo, se tradujeron al castellano los titulares de la submuestra₁.

The screenshot displays the 'Terceira Etapa' (Third Step) of the e-Termos software. The main window shows the project name 'Indização de notícias de política e economia' and the user profile 'Gerente de Projeto'. The interface includes a menu bar with options like 'Projeto', 'Recado', 'Mail', 'Lista de Termos', 'StopList', and 'Extração Automática'. A central panel displays a list of terms under the heading 'Lista: 150414_unigramas'. The list includes words such as 'siria', 'estados', 'unidos', 'china', 'partido', 'gobierno', 'dice', 'ministro', 'crisis', 'banco', and 'presidente'. A button labeled 'Visualizar Valores das Frequências' is visible above the list.

Figura 39: Etapa 3 de e-Termos, listados de términos y visualización de lista de 1 grama

4. RESULTADOS

Para evitar errores causados por elementos característicos del castellano, se reemplazaron las tildes y la letra ñ por n. Finalmente, por recomendación del sistema, también fueron eliminados aquellos términos con una frecuencia inferior a 2 en el conjunto de titulares.

Las etapas cuarta, quinta y sexta están orientadas a la creación de un vocabulario. No fueron aplicadas, ya que en esta memoria este proceso es realizado en Protégé.

Los resultados de e-Termos son analizados, distinguiendo entre términos en bruto iniciales, obtenidos tras la indización; términos en bruto finales, lo que supone la corrección de errores (símbolos y espacios en blanco) y elementos repetidos (singular y plural); términos revisados, susceptibles de ser incorporados al modelo; instancias revisadas, cuyas clases podrían convertirse en componentes del modelo; términos eliminados, elementos sin semántica. La Tabla 10 resume los datos de este primer acercamiento.

n-gramas	Términos brutos iniciales	Términos brutos finales	Términos revisados	Instancias revisadas	Términos eliminados
1	1795	1808	1171	368	287
2	295	307	112	81	114
3	1041	1041	50	50	941
4	437	437	24	40	373
5	213	213	6	20	187
6	110	110	2	14	94
7	70	70	0	5	65
TOTAL	3943	3982	1383	571	2028

Tabla 10: Resumen del análisis de términos extraídos con e-Termos

Los unigramas revisados se someten a un proceso de lematización con Grampal, reduciendo errores y facilitando la reasignación de términos posteriormente.

4.3.3. MODIFICACIÓN DE ONA

A partir de los resultados de la indización humana se estima conveniente la inclusión de cuatro nuevos elementos en la ontología ONA, que se describen a continuación.

Propiedad	Descripción	Dominio	Rango
<i>esAntecedentePara</i> (<i>isAntecedentFor</i>)	Propiedad que expresa la relación de antecedencia de una Acción sobre otra, cuando esa relación no es de causalidad ni de consecuencia. Inversa: <i>tieneAntecedente</i>	Accion	Accion
<i>esParteDe</i> (<i>isPartOf</i>)	Expresa relaciones jerárquicas de tipo partonómico (parte-todo). Inversa: <i>tieneParte</i>	---	---
<i>tieneAntecedente</i> (<i>hasAntecedent</i>)	Permite expresar la relación entre una Acción y otra que la antecede, cuando esa relación no es de causalidad ni de consecuencia. Inversa: <i>esAntecedentePara</i>	Accion	Accion
<i>tieneParte</i> (<i>isPartOf</i>)	Propiedad definida para expresar relaciones jerárquicas de tipo partonómico (todo-parte). Inversa: <i>esParteDe</i>	---	---

Tabla 11: Propiedades tipo objeto incorporadas a ONA

Estas propiedades, representadas en la Figura 40, permiten establecer relaciones entre clases cuando una es parte de otra, así como relaciones de antecedencia. En este caso, existe un vínculo temporal entre dos acciones, que es distinto al que puede darse entre una acción y sus causas y/o sus consecuencias.

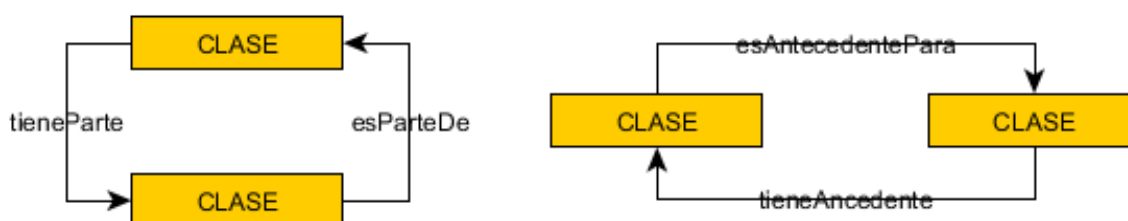


Figura 40: Relaciones partitivas y de antecedencia entre clases de ONA

El anexo B. ESPECIFICACIÓN DE ONA describe el modelo de ONA.

4.3.4. MODELADO INICIAL DE ONAPE A PARTIR DE LOS RESULTADOS DE LA INDIZACIÓN HUMANA

Teniendo en cuenta ONA y el contexto (noticias) en que aparecen los términos, se modela ONAPE. Esto implica la inclusión de elementos específicos del dominio política-economía. Por ejemplo, a partir de la expresión *Internet es uno de los instrumentos del terrorismo*, se modela *Internet* es un *Cosa* con función de *Instrumento*; *esInstrumentoDe* es una relación *Terrorismo* es una *Accion*. Como en la etapa anterior, se siguen las recomendaciones de anotación de Aguado-de-Cea et al. (2015) y Horridge (2011).

Para el modelado de ONAPE se adopta el enfoque *middle-out*, que favorece el descubrimiento de conceptos durante la construcción de la ontología (Beydoun et al., 2011), como se indica en 3.4. DESARROLLO DE LA INVESTIGACIÓN.

Middle-out es el enfoque utilizado en esta investigación, ya que se cuenta tanto con términos abstractos como en específicos. Los espacios intermedios son completados a partir del análisis de los resultados de la indización humana y del contexto en que aparecen. Así, los términos son clasificados en clases de ONA: *Acción*, *EspacioTemporal*, *Espacio Físico*, *Sujeto* y *Objeto*. Después, se definen subclases y subrelaciones. La Figura 41 muestra un ejemplo de aplicación de este enfoque: durante la indización humana se asigna la instancia *Partido de los Trabajadores del Kurdistán* y la clase *Partido político*. La clase *Organización* es heredad de ONA y *Organización política* es definida como subclase de ésta. Las relaciones *subclaseDe* e *instanciaDe* también existen en ONA, aunque no se definen explícitamente.

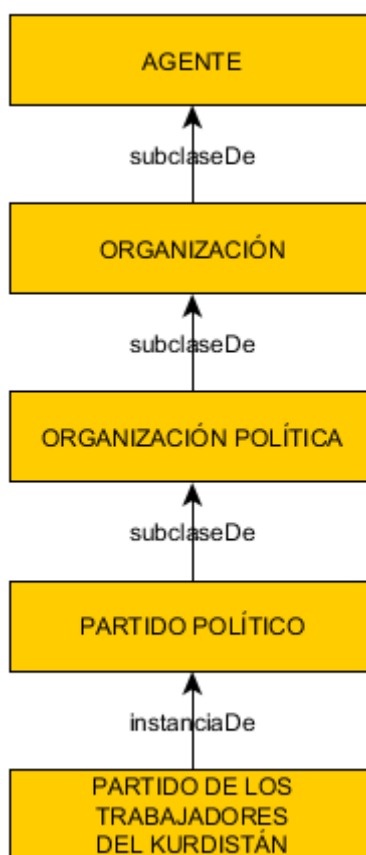


Figura 41: Aplicación del enfoque middle-out en el modelo

Para cada término extraído se analiza tanto su contexto (las noticias), como su descripción en fuentes de referencia, teniendo en cuenta la definición de los conceptos y sus relaciones. De esta forma, agentes colectivos (gobiernos, empresas, consumidores) que suelen ser asimilados con territorio (García Gutiérrez, 2014) son clasificados como Agentes, por ejemplo, Gobierno de Portugal. Por tanto, los conceptos que se incorporan a ONAPE pueden provenir de cuatro tipos de fuente:

1. **Elementos obtenidos por indización**, sean conceptos o instancias. En caso de instancias, se identifica (e incorpora al modelo) su clase;
2. **Conceptos que forman parte de la definición de aquellos asignados en indización humana**. Por ejemplo, en la Figura 42, el término *Agencia de Calificación de Créditos* es descrito como sinónimo de *Agencia de Clasificación de Créditos* y *Agencia de Calificación de Riesgos*. El primero es obtenido por indización humana, los otros aparecen en Wikipedia³⁰ y son representados en el modelo a través de propiedades de anotación;

30 https://es.wikipedia.org/wiki/Agencia_de_calificación_de_riesgos (último acceso: 15/10/2016).

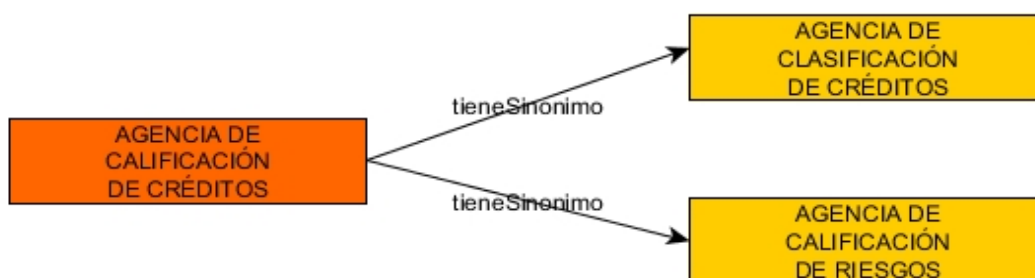


Figura 42: Relaciones de sinonimia entre términos del modelo de ontología

3. **Intrarrelaciones jerárquicas, asociativas y de equivalencia** entre términos y conceptos de vocabularios empleados para la definición de elementos de ONAPE. Por ejemplo, *Coche TR Industria automovilística*;
4. **Extrarrelaciones entre conceptos de un tesoro** con otro a través de relaciones. Por ejemplo, los elementos del tesoro EUROVOC mantienen relaciones del tipo *Has Exact Match*, *Has Narrow Match* y *Has Related Match* con otros vocabularios.

Estos procesos reflejan la complejidad de la ontología a la hora de tomar decisiones. Aun intentando modelarla de la forma más abstracta posible, con el propósito de facilitar su utilización en distintos contextos, el sistema político/económico cambia en cada país. Por ejemplo, en España se atribuye el poder legislativo al Parlamento, mientras que en otros países, este poder recae en otras instituciones, al no haber separación de poderes. En todo caso, se evalúa «qué conceptos agregar, y no [intenta no] caer en la tentación de incluir términos de forma indiscriminada» (Motz et al., 2009).

En relación con las herramientas para las definiciones de conceptos, se utilizan tesauros de ámbito europeo e internacional, generales y especializados:

- **EUROVOC**³¹, tesoro multilingüe especializado en las labores parlamentarias de la Unión Europea, que son útiles como fuentes de tipo 3 y 4;
- **GEMET** (*General Multilingual Environmental Thesaurus*)³², facilita la definición de conceptos mediante propiedades de anotación, ya que suele incluir una definición de conceptos a partir de fuentes propias oficiales, esto es, puede ser utilizado como fuente de tipo 1 y 2;
- **AGROVOC**³³, tesoro especializado en agricultura de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), ha sido actualizado a mediados de 2017. Complementa la información del resto de tesauros;

31 EUROVOC: <http://eurovoc.europa.eu/drupal/?q=es> (último acceso: 22/05/2017)

32 GEMET: <https://www.eionet.europa.eu/gemet/> (último acceso: 22/05/2017)

33 AGROVOC: <http://artemide.art.uniroma2.it:8081/agrovoc/agrovoc/en/?clang=en> (último acceso: 01/05/2017)

4. RESULTADOS

- **ECLAS** (*European Commission Libraries Catalogue Thesaurus*)³⁴, tesoro de las bibliotecas de la Comisión Europea;
- **UNBIS** (*United Nations Bibliographical Information System*)³⁵, complementa la información del resto de tesauros.

Ahora bien, los tesauros clasifican y relacionan conceptos, pero no suelen describirlos más allá de algunas notas de alcance, a excepción de GEMET. Para las definiciones de elementos de ONAPE se recurre al DLE, Wikipedia y, en menor proporción, al Diccionario Económico de Expansión y Wolters Kluwer. En el caso de Wikipedia, si no existen fuentes suficientemente acreditadas o el artículo es controvertido, se recurre a su versión en inglés.

- El **DLE**³⁶ es un diccionario de referencia en el mundo de habla hispana, actualizado permanentemente y en continua revisión. Actualmente se encuentra en su vigesimotercera edición. También ha sido útil DIRAE³⁷, un diccionario inverso basado en el DLE que, en «lugar de hallar la definición de una palabra, como en un diccionario ordinario, halla palabras buscando en su definición» (Rodríguez Alberich, 2014), lo que facilita la identificación de subclases y sinónimos;
- **Wikipedia**³⁸ es «la enciclopedia con más usuarios, autores, editores y correctores del mundo» (Cordón García, Alonso-Arévalo, Gómez Díaz y López Lucas, 2012, p. 258). La calidad media de sus artículos (exhaustividad y fiabilidad de la información) es «comparable o superior al de las enciclopedias tradicionales y, en especial, al de la Enciclopedia Británica» (Aibar, Dunajcsik, et al., 2016). «Los editores de Wikipedia utilizan mayormente fuentes altamente fiables para escribirlos» (Aibar, Minguillón, et al., 2016). Wikipedia es útil para contextualizar acciones concretas, como el escándalo financiero del LIBOR o la permanencia de países en la Unión Europea, por ejemplo. Implica el uso de DBPedia, que clasifica conceptos a partir de los contenidos recogidos en Wikipedia y mejora la categorización de sus elementos con sucesivas revisiones;
- **Diccionario Económico de Expansión y Wolters Kluwer**³⁹, especializado en economía, con más de 1600 referencias definidas por expertos/as en el dominio;
- **Enciclopedia jurídica**⁴⁰, especializada en el ámbito jurídico.

34 ECLAS: <http://ec.europa.eu/eclas/E> (último acceso: 18/07/2017)

35 UNBIS: <http://unbisnet.un.org> (último acceso: 18/07/2017)

36 DLE: <http://dle.rae.es> (último acceso: 18/07/2017)

37 DIRAE: <http://dirae.es> (último acceso: 18/07/2017)

38 WIKIPEDIA: <https://es.wikipedia.org/wiki/Wikipedia:Portada> (último acceso: 18/07/2017)

39 DICCIONARIO ECONÓMICO DE EXPANSIÓN: <http://www.expansion.com/diccionario-economico.html> (último acceso: 18/07/2017)

40 ENCICLOPEDIA JURÍDICA: <http://enciclopedia-juridica.biz14.com/> (último acceso: 18/07/2017)

4. RESULTADOS

Además, se tienen en cuenta vocabularios preexistentes, algunos de los cuales se describen en A. DESCRIPCIÓN DE VOCABULARIOS MAPEADOS, como schema.org, poblado con instancias del mundo real, DUL (*Dolce+DNS Ultralite Ontology*) y PROLES (*Political Roles Ontology*).

El siguiente apartado se centra en la modificación de ONAPE, a partir de términos extraídos por indización automática con e-Termos.

4.3.5. MODIFICACIÓN DE ONAPE A PARTIR DE LA INDIZACIÓN AUTOMÁTICA

En esta etapa, cada término es revisado y contrastado con las noticias en que aparece, con la ontología ONAPE, cuyo modelado se inicia con el proceso anterior, y con las diversas fuentes empleadas para la descripción de clases y propiedades. Los términos e instancias revisadas (Tabla 10), se clasifican teniendo en cuenta las clases de ONA: *Accion*, *Agente*, *Cosa*, *Espacios Temporales*, *Espacios Físicos*.

Sin embargo, gran parte de los resultados aportados por e-Termos no fueron incorporados a ONAPE, especialmente en el caso de trigramas. Esto es debido, fundamentalmente, a la incapacidad del programa para utilizar la lista de palabras vacías, arrojando falsos positivos. La Tabla 12 recoge las frecuencias de términos para estas cinco clases, en cada n-grama, para los términos revisados. Conviene recordar que Objeto y Sujeto son clases auxiliares, empleadas únicamente para facilitar la relación de distintos elementos en la descripción de acciones.

n-gramas	Acciones	Espacios Temporales	Espacios Físicos	Cosas	Agentes	TOTAL
1	192	19	30	217	141	599
2	22	0	11	32	46	111
3	5	0	2	22	21	50
4	8	1	1	5	10	25
5	4	0	0	2	1	7
6	1	0	0	2	0	3
7	1	0	0	2	0	3
TOTAL	234	20	42	281	220	798

Tabla 12: Frecuencias n-gramas por clase de ONA para los términos revisados.

Otros términos aparecían en más de un grama o bien fueron considerados clases siendo instancias. Es el caso de *Fondo de Rescate Permanente* (4 gramas) y *Fondo de Rescate Permanente de la Zona [Euro]* (7 gramas); *Presidente del Banco* (3 gramas) y *Presidente del Banco Central Europeo* (5 gramas). Otras instancias permitieron la definición de clases e incluso otras instancias. Un ejemplo de lo anterior es la clase definida *Estacion*. Un tipo

4. RESULTADOS

de clase definida es aquella definida a partir de sus instancias, no contemplándose la posibilidad de instancias de dicha clase más allá de las indicadas.

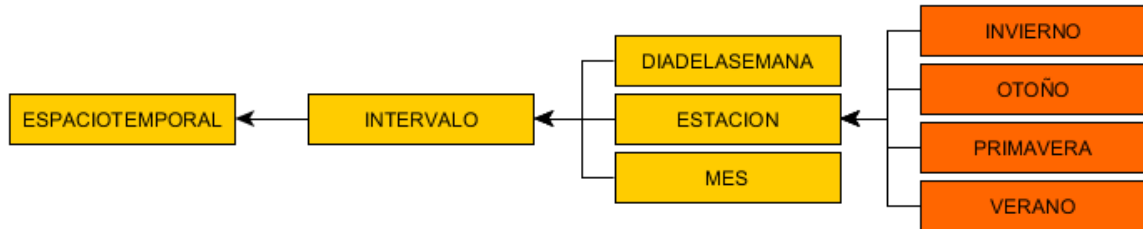


Figura 43: Ejemplo de clase definida a partir de instancias extraídas por indización automática

La Figura 43 muestra 5 clases, siendo *DiaDeLaSemana*, *Estacion* y *Mes* subclases de *Intervalo*, e *Intervalo* subclase de *EspacioTemporal*. Como instancias de *Estacion*, aparecen *Invierno*, *Otoño*, *Primavera* y *Verano*, siendo éstas todas las instancias posibles de la clase *Estacion*. En consecuencia, *Estacion* es una clase definida.

La incorporación de un elemento, por tanto, aporta no sólo una clase o propiedad, genera otros elementos necesarios para la descripción adecuada de *Acciones*, *Agentes*, *Cosas*, etc. Pongamos un ejemplo más: el del hexagrama e instancia *Embajada de Estados Unidos en Yemen* generaría 2 clases (representadas en color naranja en la Figura 44), 3 instancias (en color amarillo) y 3 propiedades: *esUn* (relación enumerativa entre clase e instancia); *representaAlEstado* y *representaAlEstadoAnte*.

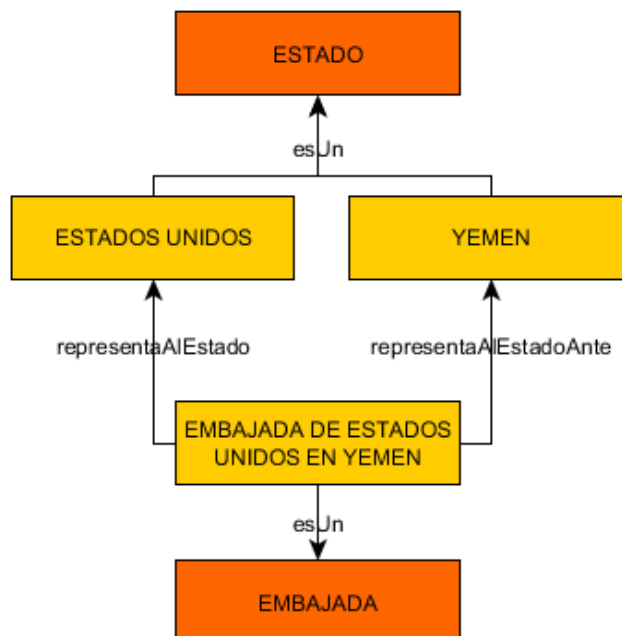


Figura 44: Clases y relaciones de ONAPE a partir de Embajada de Estados Unidos en Yemen

De forma similar al análisis de términos realizado tras la indización humana, los conceptos que finalmente se incorporan a ONA u ONAPE pueden provenir de elementos extraídos de la indización automática con e-Termos, conceptos incluidos en la definición de los aquellos extraídos por indización, intrarrelaciones jerárquicas, asociativas y de equivalencia entre términos y conceptos de los tesauros, así como extrarrelaciones con conceptos de otras herramientas utilizadas como fuente en esta investigación, como EUROVOC, GEMET, AGROVOC, ECLAS, UNBIS, el DLE, Wikipedia y, en menor proporción, el Diccionario Económico de Expansión y Wolters Kluwer. Algunas instancias fueron recogidas en el modelo para facilitar el entendimiento de sus clases. Por ejemplo: *Disputa Territorial de las Islas Senkaku*, instancia de *Accion Conflicto*⁴¹.

El tamaño de la ontología hace impracticable la descripción de todos sus elementos en esta memoria.

4.3.6. DATOS ESTADÍSTICOS DEL PROCESO

El modelo de ONA, planteado en el apartado anterior, ha sido modificado con 4 nuevas propiedades tipo dato: 7 clases, 32 propiedades tipo objeto, con sus respectivos dominios y rangos. Además, se han definido 14 propiedades inversas, 149 propiedades de anotación, de las cuales 58 son comentarios (*comment*), 27 de definición (*isDefinedBy*) y 58 son *labels*.

Los datos estadísticos del modelado de ONAPE a partir de los términos atribuidos a las noticias de la submuestra₁ son:

- **Términos iniciales:** 663, incluyendo instancias que, en general, no son recogidas en las ontologías modeladas, como se apunta en la página 36;
- **Clases:** 876, relacionadas mediante 1.009 relaciones jerárquicas de clase, 14 clases equivalentes y 27 son clases disjuntas. Algunas clases son polijerárquicas;
- **Propiedades tipo objeto (relaciones):** 65, entre las que existen 24 relaciones de jerarquía. Además, se han definido 17 propiedades inversas y 1 disjunta. También se han especificado 49 dominios y 53 rangos;
- **Propiedades tipo dato:** 10, con 4 relaciones de jerarquía, 3 dominios y 12 rangos;
- **Instancias:** 41, la mayoría incorporadas para facilitar el entendimiento de conceptos;
- **Propiedades de anotación:** 2.670, de las cuales 979 son comentarios (*comment*), 672 de definición (*isDefinedBy*), 946 son *labels* y 73 de tipo ver también (*seeAlso*).

⁴¹ Este caso es utilizado en el proceso de validación como muestra de descripción de acciones empleando el modelo planteado.

Estos datos se muestran gráficamente en la Figura 45.

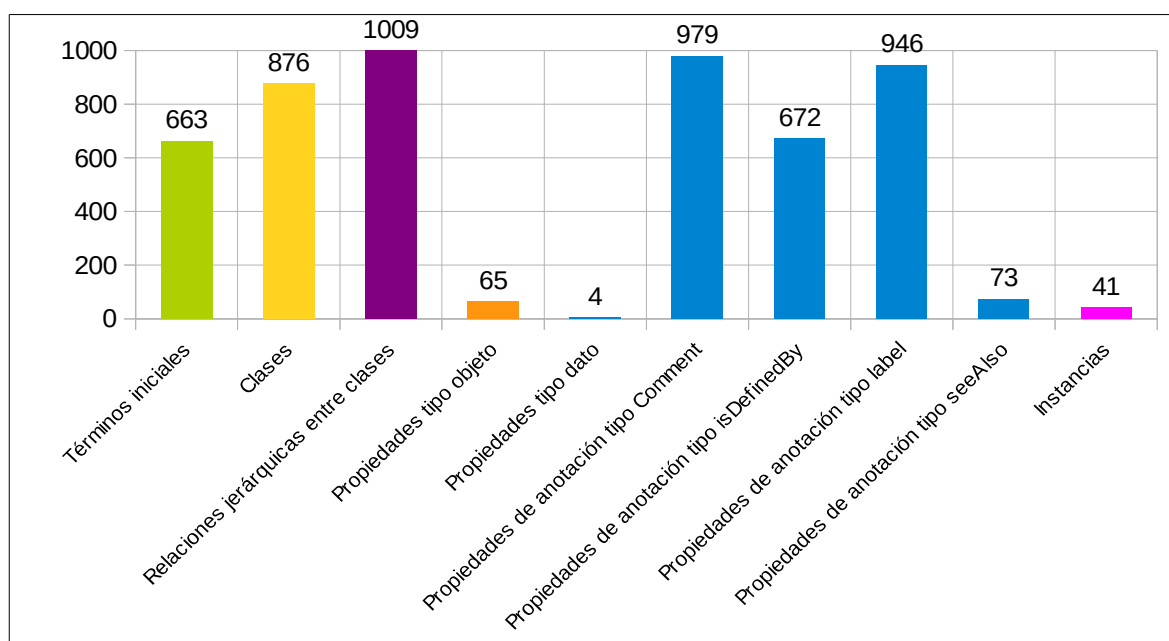


Figura 45: Datos estadísticos a partir de la indización humana

Los datos estadísticos del modelado de ONAPE a partir de los términos extraídos por indización automática de los titulares de las noticias de la submuestra₁ son:

- **Términos iniciales:** 3.982, incluyendo instancias, cuyas clases pueden ser incorporadas o no al modelo;
 - **Clases incorporadas al modelo:** 1.047, relacionadas entre sí a través de 1.175 relaciones jerárquicas de clase, 9 son clases equivalentes y 28 son clases disjuntas;
 - **Propiedades tipo objeto (relaciones):** 69, entre las que existen 36 relaciones de jerarquía. Además, se han definido 20 propiedades inversas y 1 disjunta. También se han indicado 66 dominios y 69 rangos;
 - **Propiedades tipo dato:** 14, con 4 relaciones de jerarquía, 3 dominios y 12 rangos;
 - **Instancias:** 83, incorporadas para facilitar el entendimiento de conceptos;
- Propiedades de anotación:** 3.370, de las cuales 1.207 son comentarios (*comment*), 900 definiciones (*isDefinedBy*), 1.172 son *labels* y 91 de tipo ver también (*seeAlso*).

4. RESULTADOS

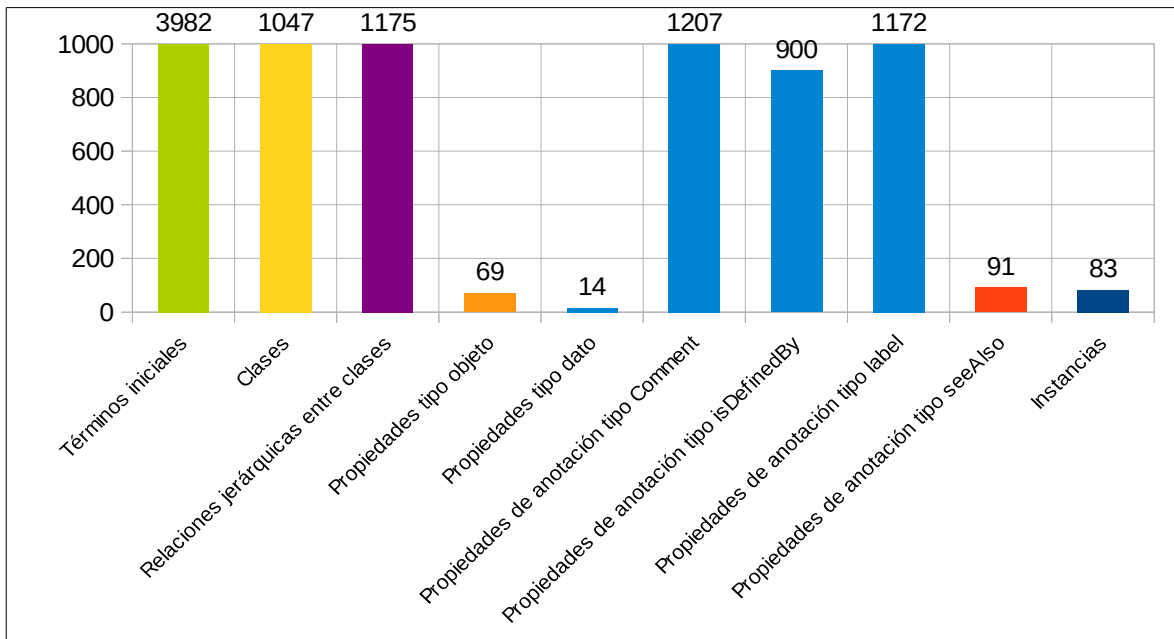


Figura 46: Datos estadísticos a partir de indexación humana + automática

La Figura 46 muestra acumulativamente cómo ONAPE es modificada con la inclusión de nuevos elementos a partir de la indexación automática.

4.4. REUTILIZACIÓN DE ELEMENTOS PREVIOS: MAPEO CON OTROS VOCABULARIOS

Una de las bases de la web semántica es la reutilización de vocabularios. Sin embargo, la integración de ontologías es compleja, especialmente cuando hay conceptos obviamente similares y no está claro cómo (Ushold y Grüninger, 1996, p. 16) establecer relaciones entre estos.

En este apartado se describen vocabularios candidatos a reutilización, tanto en ONA como en ONAPE. La Tabla 13 muestra las URL de los vocabularios disponibles en Internet y la fecha de acceso.

ONTOLOGÍA	ACRÓNIMO	ENLACE	ACCESO
<i>Basic Formal Ontology</i>	BFO	http://ifomis.uni-saarland.de/bfo/	20/05/2017
<i>Dbpedia</i>	DBPEDIA	http://dbpedia.org/ontology/	20/05/2017
<i>Schema.Org</i>	SCHEMA	http://schema.org/	20/05/2017
<i>Dolce+Dns Ultralite Ontology</i>	DUL	http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite	20/05/2017
<i>Dublin Core Metadata Initiative</i>	DCMI	http://dublincore.org/documents/2010/10/11/dcmi-type-vocabulary/	20/05/2017
<i>Conceptual Reference Model</i>	CIDOC crm	http://www.cidoc-crm.org/Version/version-6.2	20/05/2017
<i>Upper Mapping and Binding Exchange Layer</i>	UMBEL	http://umbel.org/	20/05/2017
<i>The Event Ontology</i>	EVENT	http://motools.sourceforge.net/event/event.122.html	20/05/2017
<i>Ontology for Linking Open Descriptions of Events</i>	LODE	http://linkedevents.org/ontology/	20/05/2017
<i>F-Model of Events</i>	F-MODEL	https://www.uni-koblenz-landau.de/de/koblenz/fb4/AGStaab/Research/ontologies/events/index_html	20/05/2017
<i>Friend of A Friend</i>	FOAF	http://xmlns.com/foaf/spec/	20/05/2017
<i>The Person Core Vocabulary</i>	PERSON	http://www.w3.org/ns/person	20/05/2017
<i>Vcard Ontology</i>	vCARD	http://www.w3.org/TR/vcard-rdf/	02/10/2015
<i>Vocabulary for Biographical Information</i>	BIO	http://vocab.org/bio/	20/05/2017
<i>Vocabulary for Poderopedia</i>	PODER	http://dev.poderopedia.com/documentation/index	20/05/2017
<i>The Timeline Ontology</i>	TL	http://motools.sourceforge.net/timeline/timeline.html	20/05/2017
<i>The Time Ontology</i>	TIME	http://www.w3.org/TR/owl-time/	20/05/2017
<i>Time Event Ontology</i>	TEO	http://informatics.mayo.edu/CNTRO/index.php/TEO	20/05/2017
<i>Wgs84 Geo Positioning</i>	GEO	http://www.w3.org/2003/01/geo/	20/05/2017
<i>Cartcoordont</i>	CART	http://purl.org/net/cartCoord	02/10/2015
<i>The Places Ontology</i>	PLACES	http://linda.epu.ntua.gr/vocabulary/2005/the-places-ontology/	20/05/2017
<i>Fao Geopolitical Ontology</i>	GEOP	http://www.fao.org/countryprofiles/geoinfo/en/	20/05/2017
<i>BBC Core Concep Ontology</i>	BBCCORE	http://www.bbc.co.uk/ontologies/coreconcepts	20/05/2017
<i>News Storyline Ontology</i>	NSL	http://www.bbc.co.uk/ontologies/storyline	20/05/2017
<i>BBC Politics Ontology</i>	POLITICS	http://www.bbc.co.uk/ontologies/politics	20/05/2017
<i>Political Roles Ontology</i>	PROLES	http://www.essepuntato.it/lode/http://www.essepuntato.it/2013/10/politicalroles	20/05/2017
<i>Politics Ontology for web Entity Retrieval</i>	POWER	http://dmir.inesc-id.pt/project/POWER-PT_01_in_English	06/10/2015
<i>Simple News and Press Ontologies</i>	SNAP	http://data.press.net/ontology/	20/05/2017

Tabla 13: Acrónimos, URL y último acceso a las ontologías preseleccionadas

4. RESULTADOS

La selección de ontologías requiere de la consulta de múltiples fuentes (Stuart, 2016, p. 83) como son: el repositorio *Linked Open Vocabularies* (LOV) (ver 2.2.5. REPOSITORIOS DE VOCABULARIOS: LINKED OPEN VOCABULARIES), Scopus, la Web of Science (WOS), las bibliotecas de la Universidad de Murcia, Universidade Federal de Minas Gerais (UFMG) -Brasil-, Universidade Estadual Paulista Júlio Mesquita Filho (UNESP) -Brasil-, la Universidade do Minho -Portugal-, referencias bibliográficas y sugerencias de profesores e investigadores del área. También se utiliza Google y el buscador de ontologías Swoogle⁴². Este último sigue funcionando y ha sido empleado en diversos estudios (Stuart, 2016, pp. 86-87), aunque en este caso no se hallaron resultados significativos.

Esta selección se organiza en vocabularios de alto nivel, de eventos, de agentes, de espacios temporales, de espacios físicos, de prensa escrita y del dominio político-económico. Algunos vocabularios sólo pudieron ser descritos en base a referencias bibliográficas. La Tabla 13 (página 123) muestra la lista de vocabularios accesibles.

Primero se describe cada vocabulario preseleccionado. Después se especifican los criterios de selección de vocabularios a reutilizar. A continuación se resume el proceso de mapeo entre los vocabularios y ONA. En ambas ontologías, los elementos de vocabularios reutilizados se especifican mediante la propiedad *rdfs:isDefinedBy*. En una versión posterior, los elementos serán importados directamente.

El anexo A. DESCRIPCIÓN DE VOCABULARIOS MAPEADOS describe con mayor detalle autoría, propósito y estructura (clases, relaciones, propiedades tipo dato) de los vocabularios seleccionados, así como qué herramientas reutilizan. Posteriormente se proporcionan enlaces de interés de estos vocabularios.

4.4.1. PRESELECCIÓN DE VOCABULARIOS

4.4.1.1. VOCABULARIOS GENERALES

BASIC FORMAL ONTOLOGY (BFO). BFO es un conjunto de ontologías o perspectivas de la realidad entre las que se pueden destacar dos: SNAP, para aquellas entidades persistentes (*continuants*) en el tiempo; y SPAN, para aquellos ocurrentes (*occurents*) enlazadas en el tiempo, esto es, procesos, eventos, actividades y cambios. Mientras que la primera ontología sitúa a las entidades en una región espacial (*SpatialRegion*), la segunda ubica sus ocurrentes en una región temporal (*SpaceTime*), ya que tienen una sucesión en el tiempo. Un ejemplo de ocurrente es la juventud, que no esParteDe la persona pero sí de su vida (Grenon y Smith, 2004). Aunque se presenta como una ontología que puede aplicarse a varios dominios, casi todas las ontologías que la utilizan son del ámbito científico biológico, médico, etc.

42 <http://swoogle.umbc.edu> (último acceso: 23/05/2017)

CONCEPTUAL REFERENCE MODEL (CIDOC CRM). CIDOC es una de las ontologías más completas y complejas en cuanto a la definición de conceptos y *constituye una “ontología” de la información relativa al patrimonio cultural (...). Su presentación se basa en el enfoque “orientado al objeto”* (Biblioteca Nacional de Francia, 2014). Pese a su relativa complejidad, muchas de sus propiedades han sido incluidas en perfiles de aplicación y para la publicación de conjuntos de datos (Stuart, 2016, p. 71).

La clase *Place* hace referencia a una extensión en un espacio, particularmente en la superficie de la Tierra, en un puro sentido físico, independientemente de la cuestión temporal o material. Los agentes son definidos a través de la clase *Actor*, que contempla dos subclases, *Group* (reuniones u organizaciones de dos o más personas que actúan en conjunto o de forma similar debido a cualquier forma de relación unificadora) y *Person* (persona real que vive o ha vivido). Un objeto, *Thing*, es un producto físico o intelectual relativamente estable. Tanto *Actor* como *Thing* son subclases de *PersistentItem*. Un evento es una *TemporalEntity*, subclase de *Period* (*TemporalEntity* > *Period* > *Event*) y engloba un cambio de estado en sistemas culturales, sociales o físicos, provocados por una serie de fenómenos físicos, culturales, tecnológicos o legales coherentes.

DBPEDIA. Se usa como base de conocimiento de Wikipedia, puesto a disposición en la web como RDF (Stuart, 2016, p. 72). La utilidad de esta ontología reside en el volumen de información con que trabaja y su actualización continua. En 2014, DBpedia contenía 685 clases, 1.079 propiedades tipo objeto, 1.600 propiedades tipo dato, 116 propiedades especializadas y establecía 47 equivalencias con otras clases y 35 propiedades.

Event es una subclase de *Thing* y comprende varias subclases, entre las que se incluye *SocietalEvent*. *Place* se refiere a una cosa inmóvil o localizaciones, y es equivalente a una clase homónima en schema.org. También contempla *Agent*, muy parecida a la definida en FOAF, una entidad que actúa. Es concebida como la superclase de *Person* y *Organisation*. El tiempo parece gestionado a través de propiedades como *startDate*. Los objetos y entidades abstractas no tienen una superclase y aparecen en la ontología, en algunos casos, al mismo nivel que las clases ya indicadas, como *CelestialBody*.

DOLCE+DnS ULTRALITE ONTOLOGY (DUL). Es el resultado de la simplificación de algunas partes de *DOLCE Lite-Plus library* y *Descriptions and Situations ontology*, con el fin de proporcionar un conjunto de conceptos de nivel superior que facilite la interoperabilidad entre ontologías de nivel medio y bajo. DUL fue desarrollada con la metodología OntoClean para asegurar su corrección y consistencia. Algunas de las cuestiones de competencia formuladas para ser respondidas por DUL son: ¿qué participa en un evento?, ¿cuál es la localización de algo?, ¿qué partes tiene ese algo?, ¿qué sigue a qué?, ¿sobre qué trata esto?, ¿qué concepto clasifica qué entidad?, ¿cuáles son las entidades en la base de aquella situación? Se plantea qué elementos son necesarios para

4. RESULTADOS

modelar una ontología y cómo DUL puede asistir en el proceso de dicho modelado mediante sus clases, relaciones y propiedades tipo dato.

DUBLIN CORE METADATA INICIATIVE (DCMI). Conjunto de metadatos para la descripción de recursos online. Este estándar es creado y mantenido por un grupo internacional e interdisciplinar de profesionales de diversas áreas. DC trabaja a dos niveles: simple, con 15 elementos básicos de descripción, y cualificado, que incluye otros tres elementos y de cualificadores.

Un *Event* es «un ocurrente no persistente basado en el tiempo». Los metadatos para un evento pueden proporcionar información descriptiva básica para el descubrimiento del propósito, localización, duración y agente responsable asociados con un evento. Entre los eventos posibles, por ejemplo, se incluyen exhibición, *webcast*, *conferencia*, *workshop*, *día abierto*, *performance*, *batalla*, *juicio*, *boda*, etc. Un *Agent* es un recurso que actúa o tiene el poder de actuar. Una *Location* es una región especial o un lugar nombrado. Un *PeriodOfTime* es un intervalo de tiempo que es nombrada o definida por unas fechas de inicio y fin. Un *PhysicalObject* es un objeto o sustancia inanimada y tridimensional.

SCHEMA.ORG. Conjunto de vocabularios que cubre entidades y relaciones para marcar páginas HTML, probablemente de los conjuntos de elementos de datos más importantes en la actualidad (Stuart, 2016, p. 74). Schema.org está pensado para la web y, además, para su uso por motores de búsqueda. Es una iniciativa para crear un vocabulario común con el que marcar datos web en formato de microdatos (Stuart, 2016, p. 74). Existe multitud de esquemas, algunos específicos para eventos, acciones, personas y organizaciones (agentes), lugares y algunos objetos concretos. Para la cuestión temporal se recurren a propiedades tipo dato. Además, define esquemas específicos para el ámbito periodístico como *Article* y *News Article (Thing > Creative Work > Article > News Article)*, útiles para definir la temática general de un artículo (Breadcrumbs) o asignar palabras clave.

UPPER MAPPING AND BINDING EXCHANGE LAYER (UMBEL). UMBEL es un vocabulario de clases y propiedades para la descripción de ontologías de dominio. Parte de tres clases (*RefConcept*, *SuperType* y *Qualifier*) y 38 propiedades (*correspondsTo*, *isAbout*, *isRelatedTo*, *relatesTo*, *isLike*, *hasMapping*, *hasCharacteristic* and *isCharacteristicOf*) y reutiliza elementos de vocabularios externos.

Entre sus elementos destacan *Event*: ocasiones, juegos, eventos deportivos, conferencias, fenómenos naturales, desastres naturales, guerras, incidentes, aniversarios, vacaciones identificadas por un nombre o momentos notables o periodos de tiempo; *Activity*: actividades en curso que son el resultado (en su mayoría) de un esfuerzo humano, llevadas a cabo por organizaciones que asisten a otras organizaciones o individuos (en cuyo caso son conocidos como servicios, medicina, derecho, impresión, consultoría o enseñanza) o

4. RESULTADOS

bien esfuerzos individuales o de grupo para el tiempo de ocio, disfrute, deportes, juegos o intereses personales; *PersonTypes* es la superclase para los seres humanos individuales. Además incluye la asignación de títulos formales, honoríficos o culturales de individuos específicos y los nombres de personas que llevan a cabo trabajos o actividades específicas. Las organizaciones se muestran a través de la clase *Organizations*, que incluye agrupaciones formales de personas, a veces con la ayuda de medios legales o acuerdos de algún tipo (por ejemplo, entidades políticas como naciones o municipios, empresas, gobiernos, etc.). Pero también incluye grupos informales y menos definidos como etnias o tribus o aquellos grupos que comparten intereses, grupos de expertos, etc. Esta dimensión también incluye la noción de grupos humanos identificables con miembros de un conjunto en cualquier momento; para el tiempo se vale de la superclase *Time* y se utiliza tanto para un dato concreto como para periodos, indicados en distintos formatos. La clase *Entities* es la más general e incluye todas las cosas tangibles y nombrables con UMBEL, aunque es *Products* la que agrupa más elementos.

VOCABULARIOS DE ACCIONES

Una ontología de acciones trata de modelar acontecimientos, hechos que tienen lugar en un tiempo y espacio. Varios vocabularios abordan este tipo de información, algunos de los cuales se estudian en Shaw et al. (2009). La utilidad de este trabajo reside en el análisis de las acciones en relación con el resto de entidades indicadas (agentes, fechas, objetos, lugares). Una de las dificultades de reutilizar una ontología de acciones u otra es el nivel de profundidad que se desea alcanzar. La otra cuestión es el tiempo, cómo modelar intervalos, acciones que se suceden a lo largo de un tiempo (procesos), eventos que cambian en el tiempo (transformaciones), la relación entre distintos eventos, etc.

THE EVENT ONTOLOGY (EVENT). Ontología desarrollada por el *Centre for Digital Music* (Universidad de Londres) compuesta por 3 clases y 18 propiedades que describen un evento. Un evento es una «clasificación arbitraria de una región espacio/temporal por un agente cognitivo» (Yves y Abdallah, 2007).

Un evento podría tener agentes que participen activamente en éste, factores pasivos, productos y una localización en un espacio/tiempo. Actualmente es la ontología de eventos más utilizada en la comunidad Linked Data (Shaw et al., 2009).

F-MODEL OF EVENTS. Describe eventos que son objeto de discusión e interpretación por los seres humanos. Para Scherp, Franz, Saathoff y Staab (2009), estos son precisamente los elementos que componen *F-Model of Events*, junto a los documentos a partir de los que se definen y las interpretaciones a las que están sujetas los eventos, que también deben ser reflejadas en la ontología. De esta forma se definen los requerimientos funcionales de la ontología, que también tiene otros no funcionales, entre los que se encuentran su modularidad y reutilización, pues ésta puede ser aplicada a cualquier dominio.

4. RESULTADOS

Frente a otras ontologías de eventos (*The Event ontology*, *Eventory*, *CIDOC CRM*, *EVentML*) las ventajas de *F-Model*, elaborada a partir de DOLCE-DnS Ultralite son, según los autores, la inclusión de los tipos de relaciones antes indicadas, así como la posibilidad de representar las diferentes interpretaciones sobre un mismo evento. Otras diferencias son que sigue un enfoque sistemático de desarrollo y orientado a patrones para estructurar la complejidad de un modelo de eventos en pequeñas partes y que además se definen una semántica formal y unos requerimientos de extensibilidad y modularidad que no tienen en cuenta las otras ontologías.

ONTOLOGY FOR LINKING OPEN DESCRIPTIONS OF EVENTS (LODE). LODE es una ontología para la descripción de eventos históricos y para el mapeo entre vocabularios sobre eventos. Se compone de una sola clase, *Event*, que reutiliza de *Event ontology* y 7 propiedades. LODE busca responder a qué ocurre, dónde, cuándo y quién está implicado, por lo que, entre otros, excluye las relaciones de parte-todo y causalidad.

Elemento	Tipo	Definición
<i>Event</i>	Clase	“Algo que ocurre”
<i>atPlace</i>	Propiedad	Lugar nombrado o especificado relativamente donde ocurre un evento
<i>atTime</i>	Propiedad	Instante abstracto o intervalo de tiempo en que ocurre un evento
<i>circa</i>	Propiedad	Intervalo de tiempo que puede ser descrito mediante fechas y horas
<i>illustrate</i>	Propiedad	Evento ilustrado mediante algo (habitualmente un objeto media)
<i>inSpace</i>	Propiedad (funcional)	Región abstracta de espacio (punto geoespacial o región) donde ocurre un evento
<i>involved</i>	Propiedad	Objeto (físico, social o mental) implicado en un evento
<i>involvedAgent</i>	Propiedad	Agente implicado en un evento

Tabla 14: Elementos de la ontología LODE

Un evento es algo que ocurre y puede ser recogido en una noticia o por un historiador. *Event* (ver Tabla 14) es una subclase de *TemporalEntity*, esto es, *algo que ocurre de una forma delimitada en el tiempo*, que es más general que la concepción que aquí se tiene de lo que es un evento (Shaw et al., 2009). En esta definición, un evento no implica un cambio de estado, ni intenta distinguir eventos desde procesos o estados.

VOCABULARIOS DE AGENTES

Se centran en la descripción individual y colectiva de personas, grupos y organizaciones así como sus relaciones, principalmente desde un punto de vista político-social (como colectivo, partido, asociación, empresa, grupo de poder, etc.).

FRIEND OF A FRIEND (FOAF). Especificación RDF que define 13 clases y 62 propiedades que relacionan un agente (persona, organización o grupo), elementos identificativos de esa persona y sus redes sociales, que pueden ser definidas a tres niveles:

4. RESULTADOS

redes de colaboración humana, amistad y asociación. Es una de las ontologías más estables de la web semántica (Stuart, 2016, p. 72) y fue revisada por última vez a comienzos de 2014. Esta especificación funciona a dos niveles: una *core*, con los elementos básicos de descripción de un agente, independientemente de tiempo y tecnología, y un segundo nivel, cuando se incorporan elementos característicos de los agentes en redes sociales. Una extensión de FOAF es *Who am I* (WAI), que incorpora las clases *Role*, *Profile* y *Context* (contexto en que se mantiene cierto perfil), y algunas características relativas a preferencias y gustos, que toma de la ontología REOC (*Recommendations Ontology*).

THE PERSON CORE VOCABULARY (PERSON). Vocabulario con clases y propiedades mínimas necesarias para describir una persona (real) como individuo, sin tener en cuenta qué roles juega en la sociedad, cómo se relaciona con otras personas y organizaciones, etc. Se compone de una única clase (*Person*), dos propiedades tipo dato (*patronymicName*, *birthName*) y seis tipo objeto (*placeOfBirth*, *placeOfDeath*, *countryOfBirth*, *countryOfDeath*, *citizenship*, *residency*). La clase *Person* de este vocabulario es una subclase de *Person*, de los vocabularios FOAF y schema.org. Fue desarrollado como parte del programa ISA de la Unión Europea.

vCARD ONTOLOGY es la versión actualizada y adaptada a RDF del clásico vocabulario vCard, para describir personas y organizaciones. Compuesto por 62 clases y 84 propiedades tipo objeto, con elementos puramente descriptivos, como fecha o lugar de nacimiento. Incorpora otros relacionados con las redes sociales que pueden establecerse entre estos agentes.

VOCABULARY FOR BIOGRAPHICAL INFORMATION (BIO). BIO, compuesta por 42 clases y 33 propiedades, recoge términos para saber más de las personas, sus antecedentes e información genealógica. Para ello facilita la descripción de personas y una serie de eventos clave interconectados a través de los que es posible definir una secuencia temporal (biográfica) en la que se sitúan personas y sus relaciones, como se muestra en la Figura 47⁴³, de Ian Davis, en 2011.

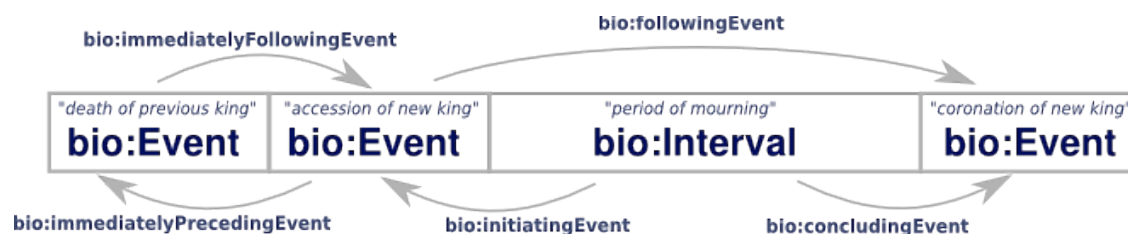


Figura 47: Biography Vocabulario Timeline

En BIO se definen propiedades de personas, de eventos (propiedades tipo dato) y relaciones entre e intra estas dos clases. Además se establece una serie de subtipos de

43 Fuente: <http://vocab.org/bio/timelines.svg> (último acceso: 23/05/2017)

4. RESULTADOS

eventos vinculados a la biografía de una persona. Mientras BIO delimita temporalmente los eventos por fechas específicas, Whois utiliza las propiedades *since* y *until*, remarcando periodos de tiempo. Además incluye otras propiedades para persona, como *career* o *honor*.

VOCABULARY FOR PODEROPEDIA (PODER). PODERVOCAB es un vocabulario para expresar información sobre personas y sus relaciones con organizaciones. Reutiliza elementos de las especificaciones FOAF y BIO. Contiene algunos elementos para el dominio político-económico, aunque parece más bien orientado al ámbito educativo.

VOCABULARIOS DE ESPACIOS TEMPORALES

En este tipo de ontologías el objetivo es establecer clases y relaciones que permitan la descripción de elementos temporales. Pueden ser más o menos descriptivas en función del propósito y vinculadas o no a otros elementos.

TIME EVENT ONTOLOGY (TEO). Ontología OWL para representar relaciones temporales entre eventos y marcas de tiempo. Se trata de una versión extendida de *Clinical Narrative Temporal Relation Ontology* (CNTRO) para su utilización en cualquier dominio. TEO se ha diseñado a partir de la ontología BFO para asegurar la interoperabilidad entre ontologías. La Figura 48⁴⁴ muestra una visión general de esta ontología.

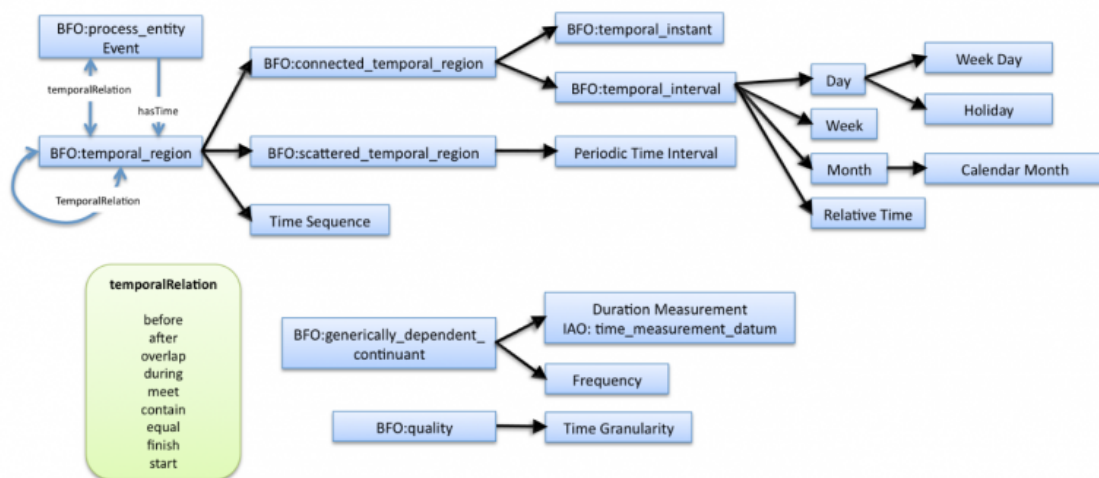


Figura 48: Vista de la ontología TEO

Un evento en TEO es equivalente a *processual_entity*, un ocurrente que sucede en un tiempo, tiene partes temporales y siempre implica y depende de alguna entidad. Se puede indicar la duración de un evento o entre eventos, así como las relaciones temporales entre estos, empleando términos como "antes" o "después de", incluso cubre periodos que se repiten. Define semánticamente regiones temporales específicas como vacaciones, días de semana, etc., así como intervalos temporales como días, semanas, meses, hoy, mañana, etc.

44 Fuente: <http://informatics.mayo.edu/CNTRO/index.php/TEO> (último acceso: 20/06/2017)

4. RESULTADOS

THE TIME ONTOLOGY (TIME). OWL-Time (antes DAML-Time) describe aspectos temporales de páginas web y sus propiedades. Su utilidad reside en establecer relaciones entre sus clases, destacando *Instant* e *Interval*, junto con la descripción de duraciones, fechas y zonas horarias. Esta ontología recoge además catorce instancias, entre las que se encuentran los días de la semana y siete unidades temporales (*unitSecond*, *unitMinute*, *unitHour*, *unitDay*, *unitweek*, *unitmonth*, *unitYear*).

THE TIMELINE ONTOLOGY (TL) también desarrollada por el *Centre for Digital Music* (Universidad de Londres) (ver Figura 49, de Raimond y Abdallah (2007)). Como *The Event Ontology*, se centra en la noción de línea temporal.

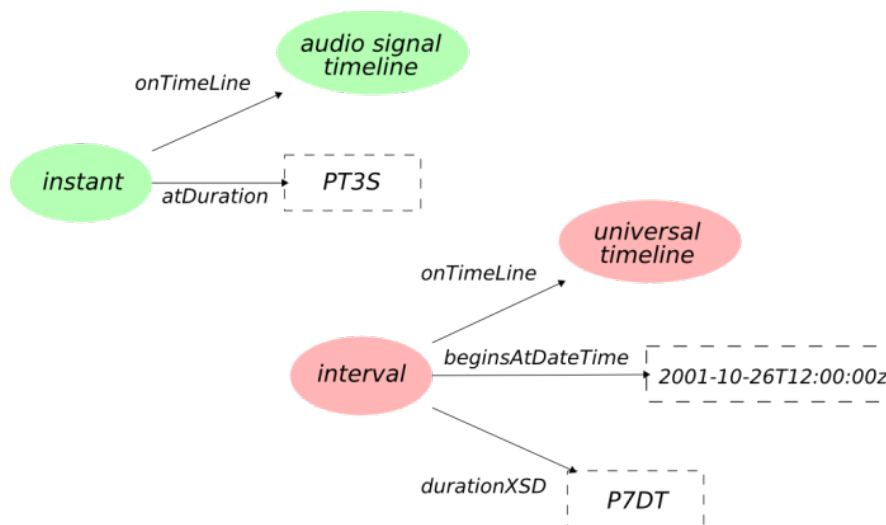


Figura 49: Clases *Instant* e *Interval* de *The Timeline Ontology*

Se compone de 22 clases y 39 propiedades que describen el inicio, fin y duración de algo, una fecha u hora concreta, la coincidencia temporal, etc.

VOCABULARIOS DE ESPACIOS FÍSICOS

Son útiles para ubicar geográficamente un acontecimiento, una persona, un objeto, etc. Pueden generar problemas cuando, por ejemplo, un territorio se encuentra en disputa o cambia de titularidad con cierta frecuencia. En estos casos resultan útiles las coordenadas geográficas. Otra solución, por ejemplo, es definir el territorio como un lugar en disputa.

CARTCOORDONT (CART). CART se utiliza para representar las coordenadas cartesianas de un espacio físico en la Tierra a partir de la clase *SpatialThing* de GEO. Se compone de tres clases, dos propiedades tipo objeto (relaciones) y tres propiedades tipo dato. La clase *Point*, subclase de *SpatialThing*, es el dominio de las tres propiedades tipo dato, *X_Coordinate*, *Y_Coordinate* y *Z_Coordinate*, una para cada una de las coordenadas geográficas.

FAO GEOPOLITICAL ONTOLOGY (GEOP). Desarrollada por la *Food and Agriculture Organization of the United Nations* (FAO) para la definición del estatus legal, autoridades y/o fronteras de cualquier territorio del mundo. La ontología recoge varias propiedades tipo dato que describen esta instancia como *agriculturalAreaTotal*, con valor 27680.0, o *nationality* con valor *español (es)*.

THE PLACES ONTOLOGY (PLACES). Ontología ligera para describir lugares de interés geográfico. Se compone de 48 clases, 4 propiedades y 1 propiedad de tipo dato. En LOV se indica que tiene muchas equivalencias con otros vocabularios. Algunas de las clases que se definen son *Peninsula*, *Ocean*, *Mountain*, *Sea*, etc. En *LinDA Transformation Engine*, donde se puede ver la descripción de sus elementos y descargar PLACES.

WGS84 GEO POSITIONING (GEO). GEO se compone de 2 clases para indicar latitud, longitud y altitud de un punto. Para ello parte de la especificación WGS84 (*World Geodetic System 1984*). En LOV, el revisor Bernard Vatant señala que, aun siendo básica, es muy utilizada en el contexto de los datos conectados y reutilizada en vocabularios, pero que, a pesar de ser publicada con el espacio de nombre de la W3C y estar firmada por Berners-Lee, no tiene estatus de oficial.

4.4.1.2. VOCABULARIOS DE PRENSA ESCRITA

Con la llegada de los medios de comunicación a Internet, muchas noticias se publican en línea y en diferentes formatos (Agarwal, Singhal y Bedi, 2012; Paepen, 2002). Por esto, no es extraño encontrar proyectos que aplican tecnologías de la web semántica (Fernández-García et al., 2006) en la producción y difusión de información de actualidad, como Fernández-García, Arias Fisteus, Sánchez y López (2012), Rubio Lacoba (2012), Fernández-García et al. (2010) y Fernández-García, Sánchez-Fernandez, Blázquez-del-Toro y Villamor-Lugo (2007). En España, por ejemplo, Rubio Lacoba y Blanco García (2010) analizan el colabulario de *El País*, una herramienta colaborativa *a medio camino entre los tesauros y las ontologías*, de ámbito general, en la que participan documentalistas y periodistas. Sin embargo, la mayoría de vocabularios no están disponibles para su descarga, por lo que se analizan artículos relacionados.

En muchos casos, el punto de partida es News Codes y, dentro de estos, casi siempre la taxonomía Subject Codes, hoy Media Topics. Tiene como aspecto positivo su rápida elaboración, pues parte de un vocabulario ya construido y actualizado con cierta frecuencia. Esta tipo de ontología es general, no se especializa en un dominio, sólo contempla relaciones de tipo jerárquico, a tres niveles, entre los conceptos y no aprovecha las noticias *per se* como fuente de términos y conceptos.

BBC CORE CONCEPT ONTOLOGY. Esta ontología permite representar elementos como personas, lugares, eventos, organizaciones y temas que tienen sentido en la *British Broadcasting Company* (en adelante, BBC). Es lo suficientemente genérico como para permitir que los/as clientes de la BBC puedan vincular sus propios conceptos, como atleta o político a un concepto concreto mediante `rdfs:subClassOf`. Esta generalidad facilita su adaptabilidad a las necesidades de cualquiera.

EPAPER. Proyecto de investigación que tiene como objetivo la creación de un suministrador central de periódicos y revistas. Para proporcionar información al lector/a, se basa en la asignación de ciertas características tanto a noticias como a los perfiles de los/as lectores/as y después busca similitudes entre ambos. Para ello, se utiliza una ontología basada en NewsCodes de la IPTC. Algunas referencias de ePaper son Maidel, Shoval, Shapira y Taieb-Maimon (2010), Shapira, Shoval, Tractinsky y Meyer (2009), Shoval et al. (2008) y Tenenboim, Shapira y Shoval (2008).

MULTIMEDIA SEMANTIC SYNDICATION FOR ENHANCED NEWS SERVICES PROJECT (MESH). MESH es un proyecto de la Unión Europea centrado en el análisis multimedia para el enriquecimiento automático de metadatos de noticias. Para ello, utiliza una ontología de dominio basada en SubjectCodes, que facilita la anotación semántica automática de noticias. Referencias sobre MESH son Kasper, Steffen y Zhang (2008a, 2008b).

NEPTUNO. Castells et al. (2006) integra una ontología ligera en RDFS para desarrollar un archivo semántico de alta calidad para Diari SEGRE, facilitando la descripción y anotación de artículos y la búsqueda de información. Neptuno combina el esquema de metadatos NewsML y Subject Codes, que es modificado a través del alineamiento con un tesoro de agencia de noticias para categorizar noticias. La ontología amplía las relaciones jerárquicas de Subject Codes a las contempladas por un tesoro: relaciones de equivalencia y asociativas, aunque sigue sin aprovechar la potencialidad de las ontologías.

NEWS ONTOLOGY EVENT MODEL (NOEM). Wang y Zhao (2012), proponen el *modelo de una ontología para entender noticias chinas*, a partir de los elementos 5W1H característicos (*what, when, where, who, why, how*) y analizan 6.000 titulares de noticias de las agencias XinHua y People. El propósito es proporcionar un vocabulario básico para anotación semántica de eventos en noticias, a tres niveles: información del evento, relaciones entre eventos y aspectos formales del documento. En este último caso, utilizan CNML (*Chinese News Markup Language*). Para la selección de conceptos se basan en ontologías de eventos previas. Como resultado, aseguran que el 85% de las noticias pueden ser descritas con su ontología. Este trabajo es, probablemente, el más parecido al desarrollado en esta memoria, aunque utiliza técnicas diferentes, y el dominio es otro.

4. RESULTADOS

THE NEWS ONTOLOGY PROJECT. Propuesta de la Unión Europea para mejorar el flujo de trabajo en agencias de noticias, desarrollando una ontología ligera en RDFS. Se basa en los IPTC Subject Codes para categorizar noticias, NITF y NewsML para la gestión de metadatos. A día de hoy parece que esta ontología no está en activo. Referencias sobre NEWS son: Fernández-García et al. (2012), Fernández-García et al. (2010), Fernández-García et al. (2007), Fernández-García et al. (2007), Fernández-García et al. (2006) y Sánchez-Fernández, Bernardi y Fuentes (2005).

NEWS STORYLINE ONTOLOGY (NSL) (Tarling y McGinnis, 2013) es una ontología ligera de eventos, en general, desarrollada por la BBC, en colaboración con *The Guardian*, Press Association y Ontoba, en 2013, para el modelado y organización de las historias de las noticias (*stories news*). De ámbito general, incluye la posibilidad de indicar quién interpreta un acontecimiento. Una historia (*story*) es una pieza individual de contenido, sea una noticia, un reportaje (*news report*) o una editorial. El sujeto que interpreta la noticia es indicado a través de la clase *Atribution*, contiene seis clases y 12 propiedades, que se muestran en la Figura 51. Podría resultar útil para el modelado inicial.

(Pena, 2012) desarrolla una ontología para describir partidos de fútbol, con el fin de apoyar la publicación de información periodística en *O Globo*. Para ello, se basa en la ontología de alto nivel UFO (*Unified Foundational Ontology*) y aplica la ontología modelada en un programa a través del que los/as periodistas pueden marcar semánticamente las noticias. Además, emplea una herramienta que facilita el uso del producto modelado por los/as periodistas, por lo que la producción de la noticia y su descripción semántica se llevan a cabo casi simultáneamente. *O Globo* emplea UFO tanto en deportes como en noticias, en general, facilitando la organización de contenidos y el reconocimiento automático de entidades (Carolo y Rocha, 2011).

PROYECTO OMNIPAPER. Desarrollado en la Unión Europea para la creación de un perfil de aplicación para el acceso a artículos periodísticos de medios europeos, utilizando Subject Codes, de IPTC. En una fase siguiente se incluyó una capa de navegación del prototipo, una versión RDF de un tesoro léxico – WordNet. La conexión a la base de metadatos era realizada a través del elemento de metadatos “omni:key_list”. Algunas referencias sobre este proyecto son Pereira y Baptista (2009), Pereira y Baptista (2004), Schranz (2005), Schranz et al. (2005), Pereira y Baptista (2004), Pereira y Baptista (2003), Yaginuma y Baptista (2004), Yaginuma y Baptista (2003a), Yaginuma y Baptista (2003b), Paepen (2002) y Yaginuma, Pereira, Ariza Ávila y Baptista (2004).

Rocha (2012) modela una ontología de dominio reutilizando UFO para la descripción de piezas informativas. Para ello, se basa en las materias por las que se categorizan artículos de varios medios de comunicación, principalmente *O Globo*, y las respuestas al Paradigma de Lasswell. Como resultado, se obtiene una modelo general que aprovecha parte del

marco teórico utilizado en esta memoria, aunque no se basa en la Gramática de Casos de Fillmore (1967) y García Gutiérrez (2014) ni técnicas de indización humana. Este trabajo no es accesible para su consulta.

SENSO KNOWLEDGE BASED ONTOLOGY analiza, etiqueta y clasifica automáticamente noticias de última hora en portugués, a partir del código fuente RSS y ecuaciones de búsqueda. Para ello, emplea una ontología que combina distintas taxonomías y ConcepNet. SENSO está compuesta por conceptos, clases generales como *Event*, *Time*, *Entity* y relaciones como *isA*, *usedFor*, *locatedAt*, *capableOf*, *madeOf*.

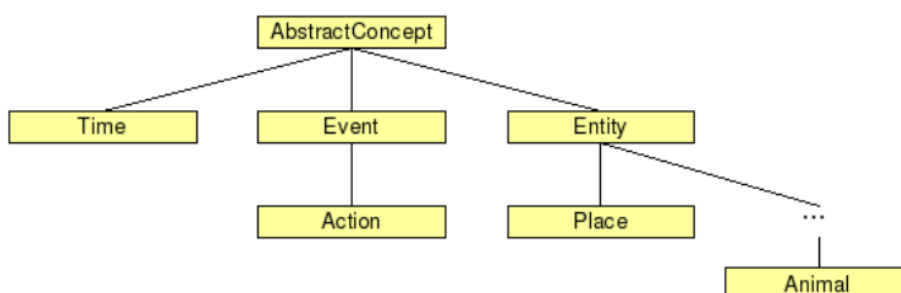


Figura 50: Elementos top level de SENSO

Este proyecto, cuyos elementos *top level* se muestran en la Figura 50, de Saias y Quaresma (2008) permite, en un entorno desarrollado *ad hoc*, hacer preguntas al sistema en lenguaje natural, obteniendo como respuesta un conjunto de noticias, gracias a la clasificación de los términos de la ecuación de búsqueda, basada en la ontología desarrollada. Otras publicaciones sobre SENSO son Saias y Quaresma (2007, 2007).

Permite plantear ecuaciones de búsqueda en lenguaje natural, con una recuperación de información más precisa, gracias a la implicación de la ontología desarrollada. Sin embargo, no aprovecha la noticia como fuente para el modelado de esa ontología ni otros elementos como esquemas de metadatos o vocabularios.

SIMPLE NEWS AND PRESS ONTOLOGIES (SNaP ONTOLOGIES) engloba seis ontologías para describir formatos (texto, imagen y vídeo) y elementos contemplados por estos (eventos, personas, lugares, etc.). Este proyecto concibe como dominio la representación del mundo real, pues esto representan las noticias (McGinnis, Wilton, Harman y O'Donovan, 2011) y distingue entre entidades simples, que denomina *stuff*, y entidades complejas, *events*. Los contenidos (*assets*) y las entidades se vinculan mediante etiquetas (*tags*). Las seis ontologías en que se divide SNaP ontologies son: Asset ontology (pna), Classification ontology (pnc), Stuff ontology (pns), Event ontology (pne), Tag ontology (pnt) e Identifier ontology (pni).

4. RESULTADOS

SMARTPUSH PROJECT es un proyecto de categorización de noticias mediante el uso de herramientas semiautomáticas y reutilización de vocabularios, generando una ontología que cubre cinco dimensiones: Materia, Ubicación, Industria, Empresa y Prioridad. Incluye, además, una herramienta (*Content Provider's Tool*) a través de la que el/la periodista puede añadir metadatos. Smartpush Project (Jokela et al., 2001) combina metadatos, una ontología predefinida, anotación semántica manual, datos de navegación y consumo de información de noticias para proporcionar información personalizada al/a la lector/a de información periodística. El objetivo es proporcionar información de interés combinando el proceso anterior con el perfil del/de la lector/a, que se analiza contra las noticias categorizadas. De esta forma el/la lector/a accede a un servicio personalizado de noticias (Jokela et al., 2001). Este proyecto se centra en el dominio financiero.

THE WORLD NEWS ONTOLOGY (WNO) es una ontología basada en NewsCodes para la anotación semántica de noticias de ámbito general y su recuperación mediante una palabras clave. WNO forma parte del proyecto WNF (*World News Finder*), para la descripción y recuperación de contenidos a partir de metadatos. A los temas seleccionados en NewsCodes (por ejemplo, accidentes y desastres, política o cuestión medioambiental), se añadieron algunos atributos. Posteriormente, se empleó GATE para el marcado automático de una muestra de noticias de agencia: temas, personas y lugares. Después se desarrolló una interfaz gráfica para la recuperación de información, permitiendo que el/la lector/a utilice también elementos de la ontología. Algunas referencias sobre esta proyecto son Kallipolitis et al. (2012), Kallipolitis et al. (2007) y Martínez-Méndez (2004).

Para el modelado de WNO se emplea la extracción automática de información, basada en PLN. Se distingue de ONA y ONAPE tanto en el dominio como en las técnicas, ya que no se reutilizan vocabularios previos, a excepción de NewsCodes. Tampoco se hace uso de la indización humana, salvo en el marcado semántico manual para obtener un *gold standard* con el que comparar los resultados del producto, ni se utilizan las noticias como fuente de elementos de la ontología.

TRONCY (2008) construye una ontología con la que enriquecer la semántica aportada por los metadatos de noticias mediante la reutilización de vocabularios y esquemas de metadatos. Primero modela una ontología a partir del esquema de metadatos (NAR) y la taxonomía News Codes, después establece vínculos con estándares multimedia, como MPEG, a continuación formaliza IPTC NewsCodes en SKOS y, finalmente, convierte los metadatos de noticias en RDF. Añade un paso más, por el que aplican técnicas de PLN a las noticias, y análisis visual de fotos y vídeos para extraer más semántica. El resultado es, además, validado a partir de un entorno de pruebas. A diferencia de la propuesta de esta tesis, no se centra en un dominio concreto ni utiliza técnicas de indización, humana y automática, para el modelado.

4.4.1.3. ONTOLOGÍAS DE POLÍTICA-ECONOMÍA

BBC POLITICS ONTOLOGY (POLITICS). Desarrollada por la BBC, describe un modelo para políticos, especialmente en términos de gobierno y elecciones locales. Se trata de una especificación en evolución, desarrollada para las elecciones europeas y locales de Reino Unido de mayo de 2014.

CORRUPTECA⁴⁵ o *Biblioteca Internacional da Corrupção* (Mesquita, Moisés y Eldasi, 2015) es un producto del grupo de investigación NUPPs (*Núcleo de Pesquisas em Políticas Públicas*), de la Universidad de São Paulo (USP). Almacena más de 100000 documentos, de cuatro ámbitos: científico, jurídico, legislativo y noticias. En este último caso, la fuente es el periódico *O Estado de S. Paulo* o *Estadão*, uno de los principales y más antiguos periódicos impresos de Brasil (Mesquita et al., 2015). La Ontología Computacional da Corrupção, utilizada para la descripción y recuperación de documentos, ha sido elaborada a partir de la colaboración interdisciplinar de politólogos/as, juristas, historiadores/as, periodistas y archiveros/as. No está disponible para su acceso y reutilización. Se cita por compartir dominio y su potencialidad.

ESSGLOBAL VOCABULARY (ESSGLOBAL) es un vocabulario desarrollado por el grupo ESSGlobal de la Red Intercontinental para la promoción de Economía Social y Solidaria (RIPESS) para la Economía Social y Solidaria (ESS), el único en LOV con términos específicos del ámbito económico, junto a ontologías como UMBEL. Reutiliza elementos de otros vocabularios como DCTERMS, Good Relations, FOAF, vCard, y schema.org.

POLIONTO. Creada por Ortiz (2011) a partir de términos extraídos y de relaciones de ontologías, esencialmente de la taxonomía government.owl, de la Universidad de Stanford, y reingeniería de ontologías. Después, utilizando técnicas de PLN se analiza un corpus de noticias de política portuguesa. No está disponible para descarga, ya que el foco del trabajo no es tanto la ontología como las técnicas por las que se modela, cuya descripción excede el propósito de esta memoria.

POLITICAL ROLES ONTOLOGY (PROLES) es una ontología basada en *Publishing Roles Ontology* (PRO), *Provenance Ontology* (PROV-O) y la ontología N-ary Participation. Permite representar atributos de los roles políticos y su posible relación con eventos y documentos de los que se extrae la información (Daquino, Peroni, Tomasi y Vitali, 2014).

A diferencia del resto de vocabularios, parte de un marco archivístico: las ISAAR(CPF) (*International Standard Archival Authority Record for Corporate Bodies, Persons and Families*) y EAC(CPF) (*Encoded Archival Context (CPF)*), adaptación de la primera a

⁴⁵ <http://corrupteca.nupps.usp.br> (último acceso: 28/04/2016)

XML, se centra en el dominio específico de política y reutiliza diversos vocabularios, como FOAF y TIME. Por otro lado, tiene en cuenta dos aspectos esenciales, que influyen en la concepción de un acontecimiento: 1) la fuente de los hechos, es decir, quién narra el evento que se describe; 2) el mismo sujeto puede participar con roles diferentes en distintos acontecimientos. Las diferencias con respecto al modelo planteado en esta memoria se centran en los procedimientos de obtención de los elementos del modelo: no utilizan técnicas de indización, usan otros vocabularios distintos y, en la perspectiva de análisis, PROLES se centra en los roles de los sujetos. Actualmente PROLES se encuentra en fase de desarrollo, teniendo como propósito su alineamiento con EAC(CPF).

POLITICS ONTOLOGY FOR WEB ENTITY RETRIEVAL (POWER). Ontología de dominio político que define sujetos políticos y organizaciones políticas, sus roles en el ámbito político y las relaciones e interacciones entre estas entidades (Moreira, Batista, Carvalho, Couto y Silva, 2011; Moreira et al., 2013). El objetivo es describir los diferentes sistemas electorales a nivel local, nacional y supranacional en el marco del proyecto REACTIONS, de periodismo computacional.

Teniendo en cuenta los elementos instanciados (políticos, asociaciones políticas, nombres de entidades políticas y mandatos en instituciones políticas) está orientada a entidades políticas (sujetos) y mandatos que en procesos políticos en general. Esta ontología fue instanciada automáticamente tomando como fuente páginas web con información estructurada, como se indica en la especificación⁴⁶. Para esta memoria, POWER es reutilizada, por compartir algunas entidades básicas (por ejemplo, Organization), por un lado, y por ser de dominio específico, por otro.

Similar a esta ontología es *Contemporaneous Role Ontology-based Expanded Query Search – Implementation and Evaluation* (CROEQS), de Vandamme, Deleu, Wauters, Vermeulen y De Turck (2009).

46 <http://dmir.inesc-id.pt/power/#sec-wherefrom> (último acceso: 20/05/2017)

4.4.2. SELECCIÓN DE VOCABULARIOS

Para la selección de ontologías se han tenido en cuenta diversos criterios con el objetivo de identificar las ontologías que mejor podrían ajustarse a los requerimientos del modelo. Para ello se tienen en cuenta estos criterios:

1. **Adecuación a la gramática de casos**, esto es, tener en cuenta cómo los vocabularios se adaptan al modelo.
2. **Difusión**. Se valora el número de vocabularios de LOV que reutilizan los vocabularios anteriores, aunque no todos aparecen en este repositorio, por lo que LOV es una buena medida, pero no determinante. En estos casos, si es posible, se tiene en cuenta la información proporcionada en el sitio web del vocabulario.
3. **Tamaño**: número de elementos que componen el vocabulario. Parece lógico asumir que, a mayor cantidad de elementos, menor necesidad de definir nuevas clases, relaciones, atributos, etc.
4. **Nivel de abstracción**. El modelo que se plantea intenta modelar una realidad compleja. Por este motivo, el nivel de abstracción del vocabulario a reutilizar no debe añadir más complejidad. Se tiene en cuenta el número de elementos que componen dicho vocabulario y la complejidad de sus relaciones.
5. **Otros aspectos**, como la disponibilidad para descarga. Se han desechado las ontologías no disponibles para su reutilización (difusión = 0), porque su uso está restringido a un ámbito interno, porque no han seguido desarrollándose, o bien porque el dominio principal del vocabulario o el ámbito geográfico que abarca no se ajusta al analizado.

La Tabla 15 muestra los vocabularios preseleccionados y la aplicación de los criterios anteriores. Los vocabularios seleccionados (y descritos posteriormente) se somborean.

Estos vocabularios son descritos con detalle en el anexo A. DESCRIPCIÓN DE VOCABULARIOS MAPEADOS. En este anexo se muestran también dos tablas sobre herramientas reutilizadas por los vocabularios seleccionados (Tabla 60), así como una serie de enlaces de interés relacionados (Tabla 61).

En el próximo apartado se muestra cómo se alinea ONA con los vocabularios analizados en este bloque.

4. RESULTADOS

	Adecuación gram. casos	Difusión	Tamaño				Nivel de abstracción	Otros
			(1)	(2)	(3)	(4)		
Ontologías de alto nivel generales								
SCHEMA.ORG	X	48	699	857	485	1540	Medio	Múltiples ámbitos
DUL	X	13	76	105	5	0	Alto	Múltiples situaciones
UMBEL	X	4	36	40	0	10	Alto	Múltiples situaciones
DCTERMS		417	34	55	12	1	Bajo	Documentos
BFO		152	35	0	0	0	Alto	Medicina
DBPEDIA		10	685	1079	1600	0	Medio	Wikipedia
CIDOC CRM		0	82	262	0	0	Alto	Cultura
Vocabularios de alto nivel: eventos								
EVENT	X	36	3	18	0	0	Bajo	Sencilla
LODE	X	11	13	15	0	0	Alto	Sencilla
F-MODEL	X	Desconocido	105	109	6	0	Alto	
Vocabularios de alto nivel: agentes								
FOAF	X	311	22	40	27	0	Bajo	Muy extendida
vCARD	X	12	62	51	33	0	Bajo	Limitado a contacto
BIO		9	57	30	3	9	Bajo	Datos biográficos
PERSON		1	7	6	2	0	Bajo	Muy sencilla
PODER		0	66	59	22	0	Bajo	Educación
Vocabularios de alto nivel: espacios temporales (tiempo)								
TIME	X	26	13	24	17	14	Bajo	Sencilla
TL	X	1	57	70	65	24	Bajo	Poco extendida
TEO	X	Desconocido	685	93	0	51	Medio	
Vocabularios de alto nivel: espacios físicos (lugares)								
GEO	X	37	2	3	0	0	Bajo	Muy sencilla
GEOP	X	2	12	6	89	312	Bajo	Sencilla y útil
PLACES	X	0	48	4	1	0	Bajo	No disponible
CART	X	0	3	2	3	0	Bajo	Muy sencilla
Vocabularios de dominio medios de comunicación								
BBCORE	X	8	5	19	0	0	Bajo	Muy sencilla
SNAP PNA		1	4	11	0	0	Bajo	Dividida en pequeñas ontologías
SNAP PNC		1	2	2	0	0		
SNAP PNE		0	1	4	0	0		
SNAP PNI	X	1	3	3	0	0		
SNAP PNS		2	6	13	0	0		
SNAP PNT		2	2	3	0	0		
NSL	X	0	6	12	0	0	Bajo	Importa quién narra
Vocabularios de dominio medios de comunicación escrita								
POWER	X	Desconocida	18	19	13	41883	Bajo	Ámbito portugués
PROLES	X	Desconocida	5	10	0	10	Bajo	
POLITICS		Desconocida	5	1	0	0	Bajo	Elecciones
ESSGLOBAL		0	15	25	0	0	Bajo	Economía Social

Leyenda: (1) Clases; (2) Propiedad tipo objeto (relaciones); (3) Propiedades tipo dato; (4) instancias

Tabla 15: Criterios de selección de ontologías para su reutilización

4.4.3. MAPEO DE VOCABULARIOS

En este apartado se describe el mapeo o alineamiento de los vocabularios seleccionados en 4.4.2. SELECCIÓN DE VOCABULARIOS. Para ello, se parte de los tipos de relación entre ontologías y las categorías básicas de los patrones de correspondencia entre éstas de Shvaiko y Euzenat (2005) y Scharffe et al. (2014), respectivamente, recogidas en la Tabla 7. El mapeo que se muestra se realiza a nivel de ONA, aunque la propiedad *rdfs:isDefinedBy*, que utiliza los vocabularios aquí analizados, entre otros, como fuente de definición, también es empleada en ONAPE.

Cuando se integran distintas ontologías, pueden aparecer distintas incompatibilidades (*mismatches*) (Bruijn et al., 2006). Habitualmente son con respecto a la ontología, en el significado o codificación de los conceptos en diferentes ontologías; en relación con el lenguaje, centrado en el formalismo utilizado para modelar las ontologías (Scharffe et al., 2014). Esta investigación se centra en el primer tipo. La Figura 51 presenta los tipos de correspondencias entre las clases de ONA y las de los vocabularios analizados.

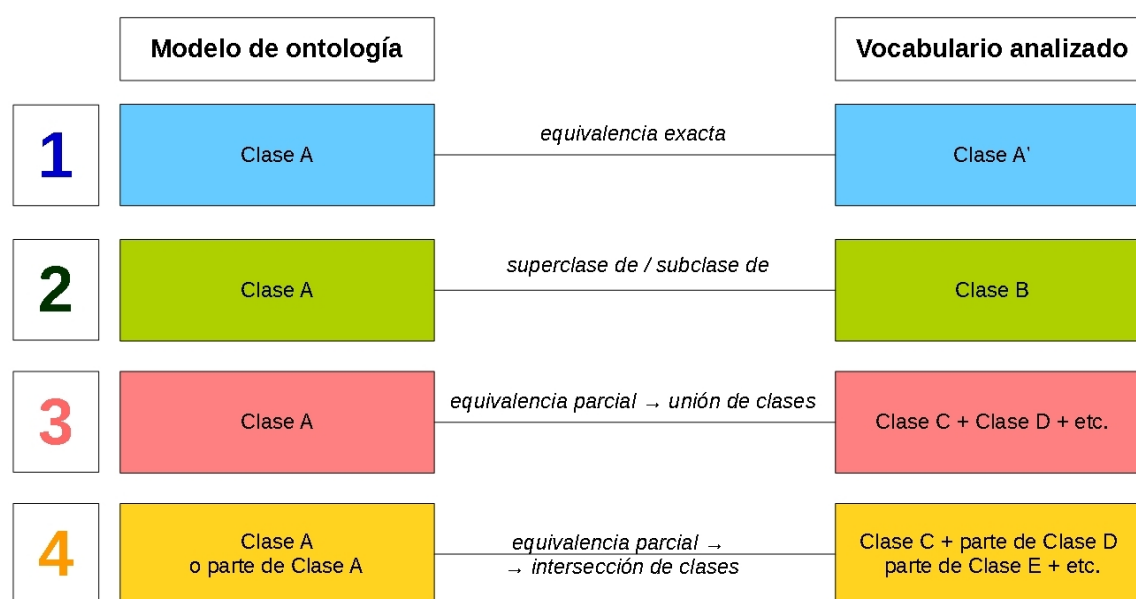


Figura 51: Mapeo de elementos del modelo de ontología y vocabularios analizados

Los sets de datos no deben ser reutilizados en su conjunto, más bien «deben ser tratados como un bufé del cual clases y propiedades son seleccionadas o referenciadas individualmente» (Stuart, 2016, p. 80). En este sentido, se analiza cada elemento de ONAPE, identificándose los tipos de correspondencia de la Figura 51, y que se describen de la siguiente forma:

4. RESULTADOS

- 1. Correspondencia tipo 1:** se da entre clases que representan el mismo concepto. Por ejemplo, entre *Accion*, del modelo, e *Event*, de EVENT, se establece una *equivalencia exacta*. Conviene recordar que, en el modelo planteado, *Accion* y *Evento* son sinónimos, esto es, representan un mismo concepto.

Un ejemplo de este tipo de correspondencia es la que se da entre *Sujeto*, del modelo, y *Agente*, de DUL (ver la Figura 52). En ONA la función agentiva la tiene el *Sujeto*, como elemento auxiliar que se relaciona con otras clases para la ejecución de la *Accion*. *Agente* tiene que ver con personas, grupos y organizaciones, en general, participando en la ejecución o recepción de la *Accion*.

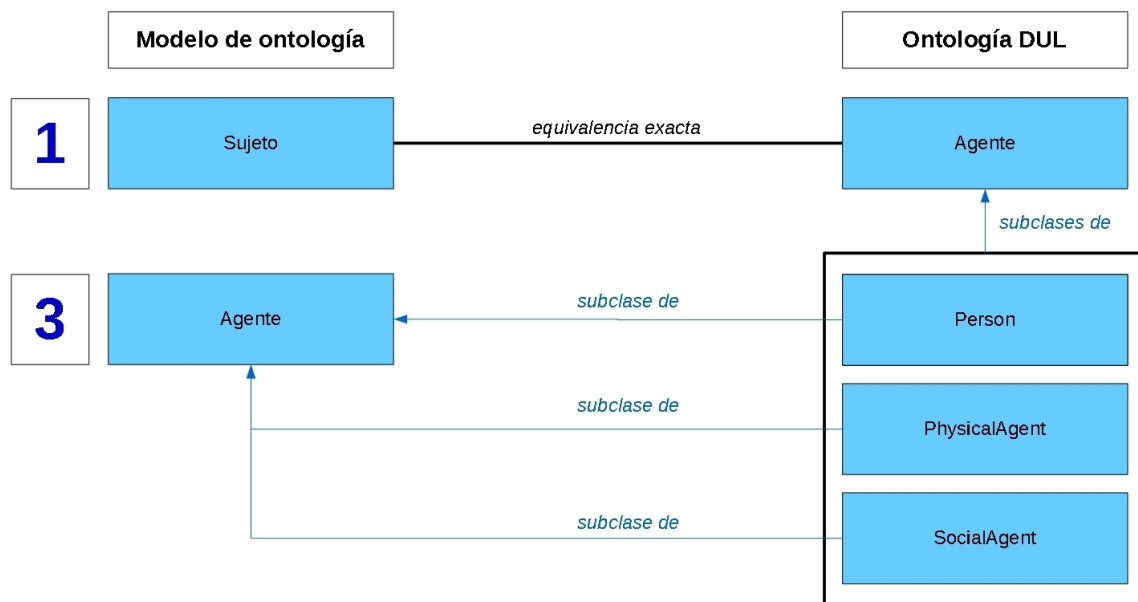


Figura 52: Correspondencias tipos 1 y 3 de Sujeto y Agente en el modelo y DUL

Esto implica la existencia de correspondencias de tipo 1 (equivalencia exacta) entre *Sujeto* (en el modelo) y *Agent* (en DUL), en el sentido de *personas, organizaciones, etc. que ejecutan una acción*, y del tipo 3 entre *Agente* (en el modelo) y *Person, PhysicalAgent* y *SocialAgent* (en DUL) (ver Figura 52), o de forma similar, entre *Agente* (en el modelo) y *Group, Person* y *Organization* (en FOAF).

- 2. Correspondencia tipo 2:** las clases del vocabulario analizado pueden ser importadas al modelo como subclases o superclases, dependiendo de cada caso. Éste puede importar nuevas subclases, pues las importadas no cubren su extensión (2+); o superclases, cuando su extensión es mayor que la clase del modelo (2-). Esta correspondencia es la que existe entre *Cosa*, en ONA, y *Document OnlineAccount* y *Project*, en FOAF. Estos tres elementos son subclases específicas de *Cosa*, pero no cubren por completo su extensión (ver Figura 53).

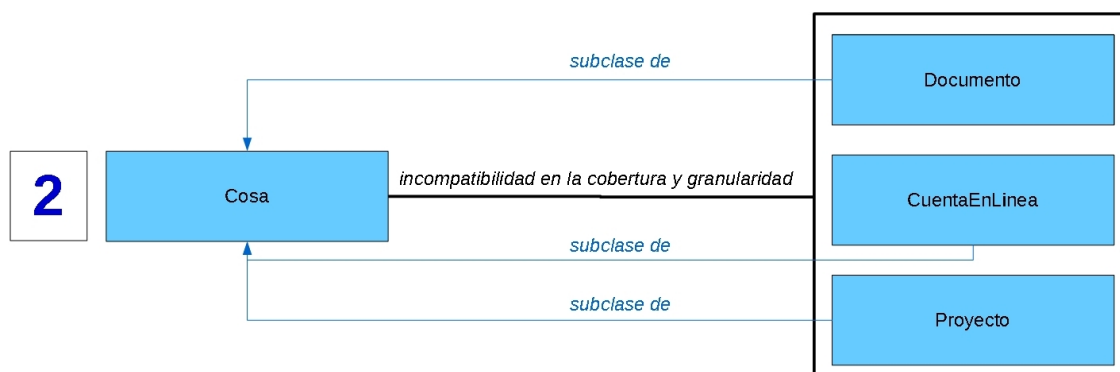


Figura 53: Correspondencia de tipo 2 entre la clase Cosa del modelo y clases de FOAF

Este tipo de correspondencia también incluye el caso contrario: que la clase del modelo sea subclase de alguna de las clases de los vocabularios analizados. Se podría hablar de una incompatibilidad (*mismatch*) derivada de las diferencias en la parte de dominio cubierto (*model coverage*) (Scharffe et al., 2014).

Un ejemplo de este tipo de correspondencia es la que puede darse entre POWER y ONA. POWER fue poblada con individuos de la política portuguesa y sólo dos tipos de entidades políticas: organizaciones y políticos. Esto excluye cualquier individuo que no se dedique profesionalmente a la política (por ejemplo, un activista) y grupos políticos que no sean asociaciones (refugiados, homosexuales, mujeres, etc.). ONA no cubre solamente el ámbito político, también el económico. Por ello, no es coherente considerar *Organization* como subclase de *Political entity*. Sin embargo, algunas clases sí pueden ser útiles, como la especificación del tipo de poder político (*executive, legislative, moderator*), la forma de alcanzar un mandato (*election, appointment, inherency*) o el alcance de una organización (*international, european, national, regional*).

- 3. Correspondencia tipo 3:** la unión de las clases del vocabulario equivale parcialmente a la clase del modelo, y pueden ser importadas como subclases del mismo, cubriendo totalmente su extensión. Por ejemplo, entre *Agente*, en el modelo, y *Group, Organization* y *Person*, de FOAF, existe una correspondencia de equivalencia parcial, como se muestra en la Figura 54.

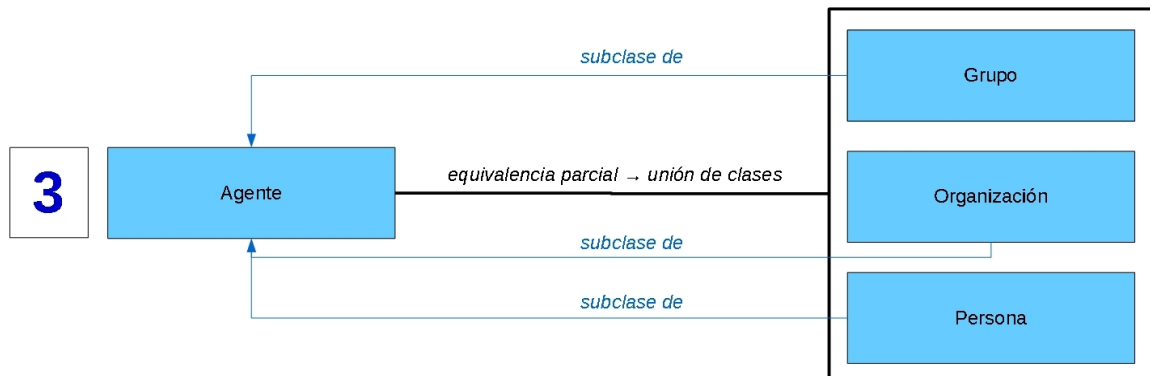


Figura 54: Equivalencia parcial entre la clase *Agente* del modelo y FOAF

Agente es equivalente a la unión de las otras tres clases. Dicho de otra forma, *Agente* no tiene más subclases, sólo *Group*, *Organization* y *Person*. Así, estos tres elementos son subclases de *Agente*.

- 4. Correspondencia tipo 4:** la intersección de las clases del vocabulario equivale parcialmente a la de ONA. Ésta podría importar estas subclases o no, pues en realidad no cubren la extensión del concepto definido en el modelo. La clase del modelo tendría otras subclases. Así se representa en la Figura 55, en la que *Cosa* es equivalente a la intersección de *Abstract* y *Object* de DUL, pues estas dos últimas tienen una extensión mayor que *Cosa* (contemplan más subclases).

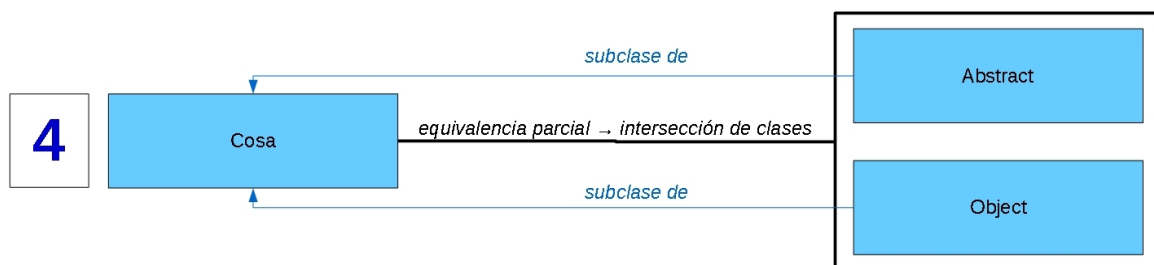


Figura 55: Correspondencia tipo 4 entre *Cosa* del modelo y *Abstract* y *Object* de DUL

La Tabla 16 muestra el mapeo de clases entre el modelo planteado y los vocabularios analizados. Las clases susceptibles de ser reutilizadas, de acuerdo a las correspondencias anteriores, aparecen sombreadas. Cuando la clase del vocabulario analizado es reutilizada, se indica con el símbolo (r).

	Accion		Agente		Cosa		EspacioFisico		EspacioTemporal		Objeto		Sujeto
BBCCORE⁴⁷	Event	1	Organization Person	3	Theme	2+	Place	2+					
DUL	Event	1	Person PhysicalAgent SocialAgent	3	Abstract Object	4	PhysicalPlace Place SpaceRegion	2+ 2+ 2+	TimeInterval	1			Agent 1
EVENT	Event	1					SpatialThing	1(r)	TemporalEntity	1(r)			Agent 1(r)
FOAF			Group Organization Person	3	Document OnlineAccount Project	2	SpatialThing	1(r)					Agent 1
GEO							SpatialThing	1					
GEOP			organisation	2+			area	2+					
TIME							SpatialThing	1(r)	TemporalEntity	1			
POWER	Election	2	PoliticalEntity RunnerList	2+ 2+	PoliticalScope PolicialPower webPage	2+							
PROLES	Event RoleInTime Situation	1(r) 1(r) 2(r)-	Group Organization Person	3(r)	Object Role	1(r) 1(r)	Place	2+	TimeInterval	2+			Agent 1(r)
SCHEMA	Accion Event	2-	Organization Person		CreativeWork Intangible Product	2+ 2+ 2+	Place	2+					Agent 1
SNAP	Event	1(r)	Organization Person	3(r)	Asset Stuff	2+ 2-	Location	2+					

Tabla 16: Clases de los vocabularios analizados para su mapeo en el modelo

47 Esta ontología reutiliza otros vocabularios como DCTERMS, pero no establece relaciones con los elementos que reutiliza, por lo que en este mapeo no se tienen en cuenta más que los elementos propios de BBCCORE

Por lo que respecta a las relaciones, de forma similar a las clases de la Tabla 16, se analizan las relaciones de ONA y las de los vocabularios analizados en el apartado anterior, indicando el tipo de correspondencia que podría darse. Para ello, se toma como referencia la Figura 51 (página 141):

- **Correspondencia tipo 1** entre dos relaciones que representan el mismo tipo de vínculo entre dos conceptos:

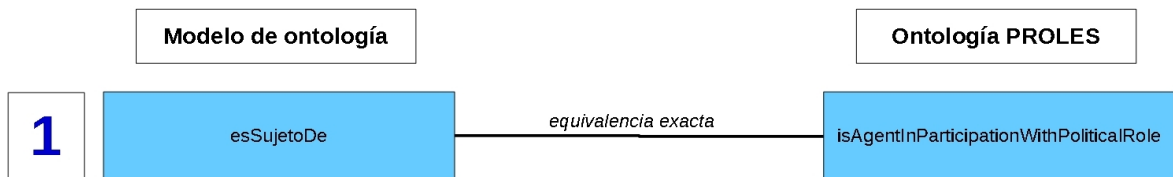


Figura 56: Correspondencia tipo 1 con una relación del modelo y una relación de PROLES

En este caso, como ya ocurría con algunas clases, no se puede decir que las propiedades sean reutilizadas tal cual, ya que, aunque cubren un universo de individuos similar, no se ajustan al propósito de ONA.

- **Correspondencia tipo 2:** las relaciones del vocabulario analizado pueden ser importadas al modelo como subrelaciones (las importadas no cubren su extensión (2-)) o superrelaciones (cuando su extensión es mayor que la relación del modelo planteado (2+)), dependiendo de cada caso.

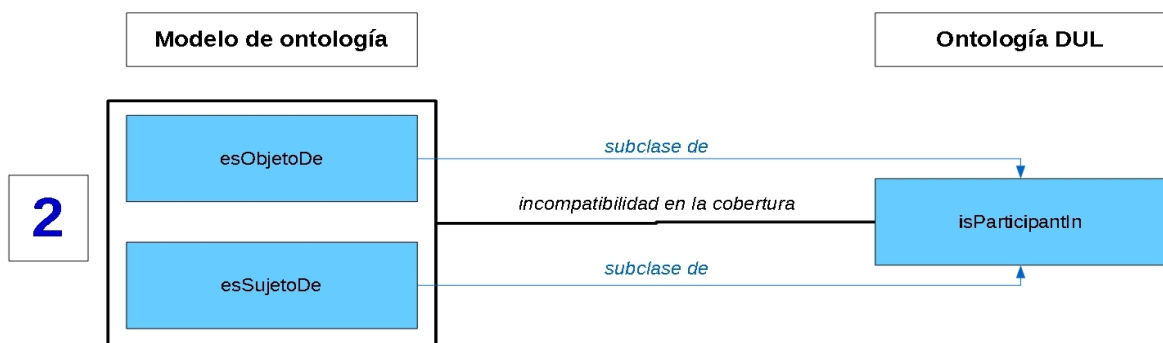


Figura 57: Correspondencia de tipo 2 entre relaciones del modelo y de la ontología DUL

En la Figura 57 se presenta una incompatibilidad entre relaciones de ONA y DUL, pues se considera que, aunque la relación *isParticipantIn* cubre la extensión de *esObjetoDe* y *esSujetoDe*, no incluye *acompañaA*. *isParticipantIn* cubre parcialmente la extensión de *esObjetoDe* y *esSujetoDe*.

De acuerdo con la especificación de Event, las relaciones *factor* y *factor_of* se pueden utilizar tanto para causas como para instrumentos, por lo que esta relación se podría concebir como la unión de *tieneCausa-tieneInstrumento* y *esCausaDe-esInstrumentoDe*, respectivamente.

4. RESULTADOS

La Tabla 17 muestra el mapeo de relaciones entre ONA y los vocabularios analizados. Las relaciones susceptibles de ser reutilizadas aparecen sombreadas. Para las siguientes relaciones no se encontraron correspondencias: *esConsecuenciaDe*, *esSituacionDe*, *esFinalidadDe*, *esModoDe*, *esReceptorDe*, *tieneSituacion*, *tieneFinalidad*, *tieneModo*, *tieneReceptor*, *tieneSubAccion*. Tampoco se han hallado correspondencias para relaciones de GEOPOLITICAL ni POWER.

Hay que tener en cuenta que muchas de las relaciones de los vocabularios analizados son específicas para el tipo de elementos que contiene cada uno. Esto es, no están pensadas, en general, para acciones, por lo que no tienen correspondencias, *a priori*, con ONA. Por ejemplo, la clase *TemporalEntity* de la ontología TIME requiere de otros elementos para completar la descripción de dicha entidad, como *DurationDescription* o *DateTimeDescription*, en el caso de intervalos de tiempo. Otra muestra es la ontología GEOP. Si bien se han establecido correspondencias para algunas de sus clases, no se ha realizado con sus relaciones por lo que no aparece referencia alguna a este vocabulario en la Tabla 17. Sin embargo, sus clases y relaciones sí pueden ser importadas posteriormente para enriquecer el modelo. De esta forma, puede ser conveniente importar las clases *area* y *organisation* y otros elementos como forma de expresar particularidades de éstas.

Existen otros elementos de los vocabularios analizados que pueden considerarse útiles para completar la ontología, especialmente en el dominio específico de política-economía. Así, en POWER, la clase *Mandate* permite indicar el tiempo que *Politician* es miembro de una *Organization*. Todos los casos son analizados posteriormente, de forma individual para su importación, o no, en el modelo. Por otro lado, se podrían incorporar otras relaciones distintas a las planteadas en ONA. Así, *hasConstituent* (de DUL), *member* (de FOAF) o *isComposedBy/isPartOf* (de POWER) se puede utilizar para indicar relaciones entre elementos, de tal forma que uno *esParteDe* otro, pero sin un sentido partonómico, por ejemplo, entre una organización y las personas que forman parte de la misma. Igual ocurre con aquellos elementos relativos a una acción en su conjunto. ONA debe incorporar una forma de indicar quién describe la acción. Esto es, lo que es una acción o no lo aporta la noticia, su redactor/a, no quien la modela. Para ello, además de la reificación (sentencias sobre sentencias), existen clases y relaciones de diversos vocabularios, como los específicos de medios de comunicación, por ejemplo *byline* (de SNAP PNA).

Este apartado da los pasos iniciales para el mapeo de vocabularios preexistentes al modelo de ontología planteado. Sin embargo, también evidencia la necesidad de un análisis pormenorizado de las correspondencias detectadas. Por otro lado, como ya se ha indicado antes, muchos de los elementos analizados no se pueden reutilizar tal cual por no ajustarse al propósito de representación de ONA. El alineamiento con el resto de vocabularios es descrito como propuesta de futuro en 5.2. TRABAJOS FUTUROS.

Propiedad	BBCORE	DUL	EVENT	FOAF	GEO	TIME	PROLES	SCHEMA	SNAP	
<i>esAcompañanteDe</i>		coparticipatesWith	2-					participant	2-	
<i>esCausaDe</i>			factor_of	2-						
<i>esEjecutanteDe</i>							isAgentInParticipationWithPoliticalRole	1		
<i>esEspacioFisicoDe</i>		isLocationOf isRegionFor	2+		location	1		location	1	
<i>esEspacioTemporalDe</i>		isTimeIntervalOf	1				isTimeIndexFor	1(r)		
<i>esInstrumentoDe</i>			factor_of	2-						
<i>esObjetoDe</i>		isParticipantIn	2-				isParticipantIn IsIncludedInParticipation isParticipantIn	1(r) 2- 2-	object	1
<i>esResultadoDe</i>			produced_in	1						
<i>esSujetoDe</i>		isParticipantIn	2- agent_in	1			isIncludedInParticipation	2(r)-	agent	1
<i>tieneAcompañante</i>		coparticipatesWith	2-							
<i>tieneCausa</i>	eventTheme	2+								
<i>tieneConsecuencia</i>								result	2-	
<i>tieneEjecutante</i>	eventOrganisation eventPerson	2+					includesAgentWithPoliticalRole	1		
<i>tieneEspacioFisico</i>	eventPlace	1	hasLocation hasRegion	2+ place	1					
<i>tieneEspacioTemporal</i>			hasTimeInterval	time	1		hasDurationDescription	2+ atTime	1(r)	
<i>tieneInstrumento</i>								instrument	1	
<i>tieneObjeto</i>			hasParticipant	2-			hasParticipant	1(r)		
<i>tieneResultado</i>			product	1	made	1	made	1(r)	result	2-
<i>tieneSujeto</i>			hasParticipant	2- agent	1	knows	2+	knows participationIncludes	2(r)+ 2-	
<i>esSubAccionDe</i>			sub_event	1					subEventOf	1

Tabla 17: Relaciones de los vocabularios analizados para su mapeo en el modelo

4.5. EVALUACIÓN DE LAS ONTOLOGÍAS ONA Y ONAPE

Como se indica en el apartado 2.2.3. METODOLOGÍAS PARA CONSTRUIR Y EVALUAR ONTOLOGÍAS, la evaluación de una ontología se realiza en dos procesos: verificación y validación.

Para la verificación, Protégé contempla varios razonadores. Tanto ONA como ONAPE han sido verificadas con el razonador Hermit 1.3.8.413 sin arrojar errores.

Por lo que respecta a la validación, para analizar la capacidad de ONAPE para describir o representar el dominio política-economía, se desarrolla una validación en dos etapas, resumida gráficamente en la Figura 58.

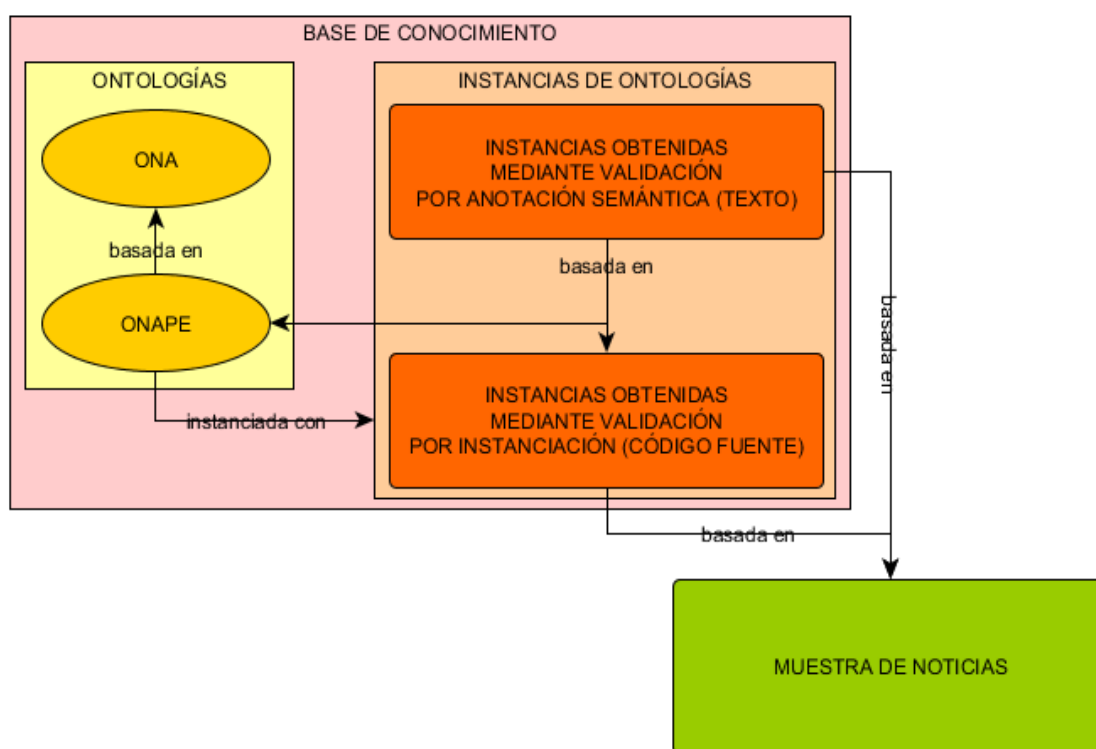


Figura 58: Representación gráfica de las dos etapas de validación de ONAPE

En esta validación, basada en datos (Hlomani, 2014), se emplean las noticias de la submuestra₂, descrita en 4.1.3. DEFINICIÓN DE SUBMUESTRAS.

En la validación mediante instanciación de ONAPE, se puebla esta ontología a partir de las palabras clave definidas por los periódicos como términos que, en su conjunto, representan el contenido de las noticias. Para ello, se analiza el código fuente de noticias, los esquemas de metadatos presentes en éste, patrones de uso de los mismos y las pautas de inclusión de las palabras clave. Este preprocesamiento es implementado en el software MetadadosHTML, descrito en el siguiente apartado.

4. RESULTADOS

En la validación mediante anotación semántica, aplicada a la misma muestra de la etapa anterior, se marca el texto de las noticias. Para ello, se utiliza la ontología ONAPE.

El dominio abarcado por ONAPE, política-economía, es complejo y amplio. Por tanto, este proceso debe ser entendido como paso inicial del aprendizaje automático (*machine learning*) para, en un futuro, obtener nuevas instancias (instanciación o poblamiento) y describir conjuntos de noticias (anotación semántica).

4.5.1. PREPROCESAMIENTO DE LA SUBMUESTRA₂ PARA LA EVALUACIÓN DE ONAPE: METADADOSHTML

Las palabras clave forman parte del conjunto de metadatos que describen y facilitan el almacenamiento y recuperación (Troncy, 2008) de piezas de información. Sin embargo, existen múltiples esquemas de metadatos para artículos de prensa. Un resumen de esquemas de metadatos para este tipo de información, como ANPA 89-3 o NITF se muestra en Baños-Moreno, Felipe, Pastor-Sánchez, Martínez-Béjar y Lima (2015). Los esquemas identificados en los códigos fuente de la muestra se pueden clasificar en:

- a) **Esquemas estándar**, como Facebook Open Graph Protocol (OGP), Dublin Core, Twitter Cards, hAtom o schema.org. También se incluyen etiquetas meta HTML relativas al contenido del documento y consideradas *nombres de metadatos estándar* (Hickson et al., 2014): *author*, *description* y *keywords*;
- b) **Esquemas *ad hoc***, no son de uso extendido y obedecen a fines específicos, como Sailthru, Bi3d, Parse.ly y DCsExt. También podría incluirse la etiqueta meta, *news_keywords*, de Google, para que los periódicos indiquen cuáles son las palabras más relevantes de una noticia (Google Inc., 2016) y que está reemplazando a *keywords* (Galfi, 2012; McGee, 2012);
- c) **Esquemas propietarios**, para la estructuración y categorización de artículos, utilizados únicamente por *The New York Times* y *The Wall Street Journal*.

Por lo que respecta a los patrones de uso de las palabras clave, aparecen en todos los periódicos de la muestra, a excepción de *Le Monde*, *Nigerian Tribune* e *Independent Online*, aunque de formas distintas. Se pueden destacar los siguientes aspectos:

- a) campos repetibles: Casi todos los medios recogen en una sola línea el conjunto de palabras clave que representa el contenido de un artículo. Sin embargo, algunos periódicos repiten la etiqueta con datos diferentes, tantas veces como palabras clave refiera;
- b) mismo propósito, información diferente. *The New York Times* muestra las palabras clave en distintas etiquetas meta, rompe el patrón de otros esquemas y no recoge la misma información, aunque el fin del esquema es el mismo;

4. RESULTADOS

- c) escasa categorización: *The New York Times* es el único diario que indica, utilizando un esquema propio, el tipo de palabras clave: *des* para descriptores; *org* para organización; *per* para persona; *geo* para lugar;
- d) ruido: se muestra en esta etiqueta información que no corresponde a este espacio.

Estas cuestiones influyen directamente en el grado de descripción de las palabras clave empleadas por los medios. Por ejemplo, *news* es la palabra más frecuente (471), seguida de *politics* (219), *breaking news* (217) y *newspaper* (194). También afectan a la capacidad de extracción de la información contenida en códigos fuente. El software MetadadosHTML, desarrollado *ad hoc* para esta investigación, tiene en cuenta estas particularidades.

Puede encontrarse más información sobre uso de esquemas de metadatos, patrones de uso de palabras clave y MetadadosHTML en Baños-Moreno, Felipe, Pastor-Sánchez, Lima y Martínez-Béjar (2017).

4.5.1.1. DESARROLLO DE METADADOSHTML

Como primer paso, se construyó una base de datos compuesta por tres Tablas (ver Tabla 18) que facilitara la relación entre sus elementos. Para esta base de datos se utilizó el software Microsoft Access. De esta forma, es posible mantener disponible en el tiempo la información extraída para su consulta.

Nombre de la tabla	Objetivo
<i>URL</i>	Almacenar las direcciones (URL) de las noticias
<i>Tags</i>	Almacenar las etiquetas meta y atributos que se desean extraer
<i>Metadados</i>	Almacena el contenido de las etiquetas meta y atributos de la Tabla Tags

Tabla 18: Tablas que componen la base de datos de MetadadosHTML

En esta etapa inicial se importaron las noticias de la muestra, almacenadas inicialmente en una hoja de cálculo, a la Tabla URL. Posteriormente y, a partir del análisis realizado con anterioridad, se registraron etiquetas meta y atributos en la Tabla Tag. Este proceso conlleva un importante esfuerzo pues, como ya se ha indicado en este apartado, diversos periódicos optaron por utilizar notaciones diferentes, sin seguir un patrón y, a veces, en formato propietario. Esta infraestructura informacional (ver Figura 59) daba inicio al desarrollo propiamente dicho del software (v1), cuyo algoritmo se basaba en el lenguaje Object Pascal.

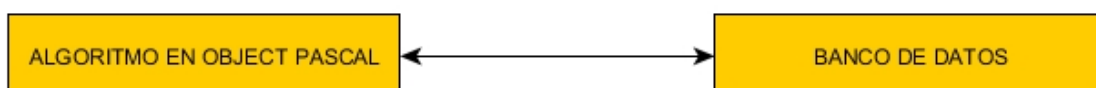


Figura 59: Infraestructura organizacional de MetadadosHTML (v1)

4. RESULTADOS

MetadadosHTML tiene tres funcionalidades:

- Acceso al código fuente de las URL contenidas en la hoja de cálculo inicial. Debido al gran número de URL, se descargaron todos los códigos HTML de las noticias en local, de forma que su análisis fuera realizado sin necesidad de acceder de nuevo a las páginas web. De esta forma se favorece la independencia de la conexión de datos y mayor velocidad de procesamiento posterior.
- Definición de las etiquetas meta y atributos de cada medio, teniendo en cuenta sus particularidades y formatos. El software analiza el documento para la extracción de los valores de dato deseados. La ausencia de patrones de notación en parte de los *corpora*, dificultó la extracción de información. La estrategia adoptada para resolver este problema fue, junto a las etiquetas meta y atributos de la Tabla Tags, la inclusión de los marcadores de inicio y fin de cada elemento. Esto sería suficiente para documentos bien formados y válidos según la notación W3C.
- Extracción del contenido de los códigos fuente, exportación a una hoja de cálculo y almacenamiento para su consulta y análisis. Tras el acceso y descarga del código fuente, comienza la extracción de metadatos. El algoritmo repite este proceso para cada página web a la que accede, que ha sido grabada anteriormente. Entonces recorre la Tabla Tags y busca el metadato. En caso de hallarlo, extrae su valor, que se encuentra entre los límites de la etiqueta. Esta información es almacenada en la Tabla Metadatos, con referencia a URL, código fuente y resto de metadatos.

La Figura 60 muestra el interfaz de MetadadosHTML (v1), desarrollada en Object Pascal.

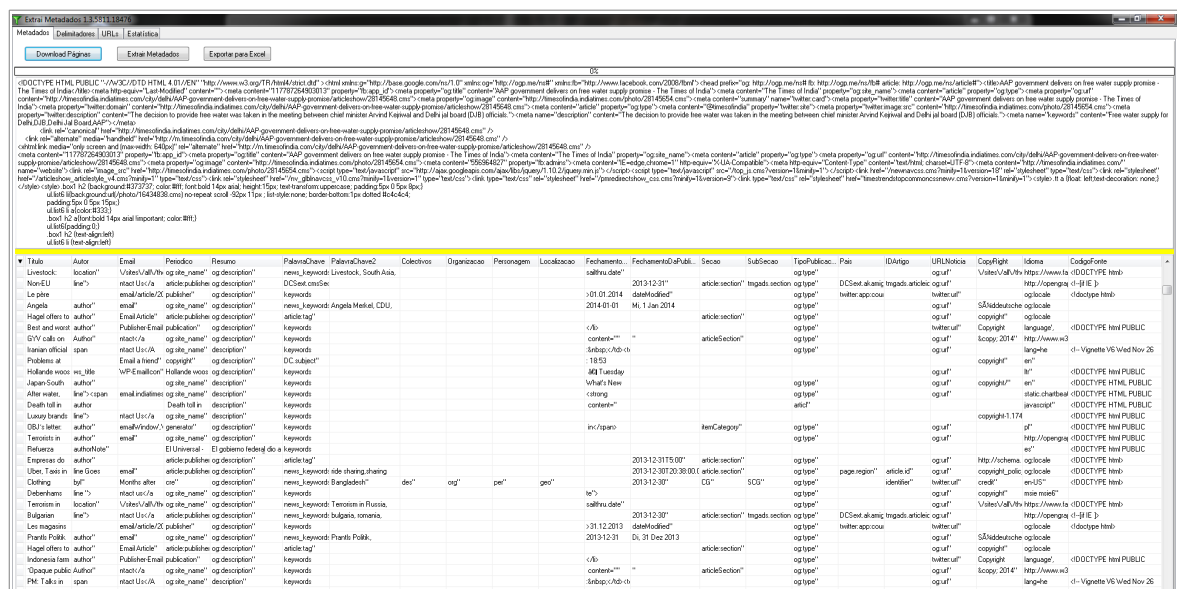


Figura 60: Interfaz de MetadadosHTML (v1)

4. RESULTADOS

Este esquema estructural no difiere de la construcción anterior (v1), modificando sólo la capa de procesamiento, esto es, el algoritmo. La lectura de las URL y la grabación de los resultados para el análisis posterior siguen usando la estructura de base de datos del modelo anterior. La interfaz de esta versión de la herramienta se muestra en la Figura 61. La extracción de los datos para su análisis se realiza directamente en la base de datos.

Posteriormente, para mejorar la extracción de metadatos en los archivos codificados en HTML, se modificó la plataforma en cuanto a la construcción de MetadadosHTML (v2). En realidad, se desarrolló un nuevo software para las operaciones de alto nivel de procesamiento de los datos, siendo reutilizados de la etapa anterior la base de datos, con sus tablas y relaciones. Este software fue desarrollado en Java y tiene como aspecto más relevante la adopción de jsoup*, un *parser* HTML⁴⁸ ampliamente utilizado en la comunidad de desarrolladores. La ventaja de este *parser* frente a *Object Pascal* es el mayor grado de especificidad en el análisis del código HTML y sus variantes, así como una mejor detección y tratamiento de errores. De esta forma, se reduce el número de situaciones indeseadas en el procesamiento de los identificadores de metadatos.

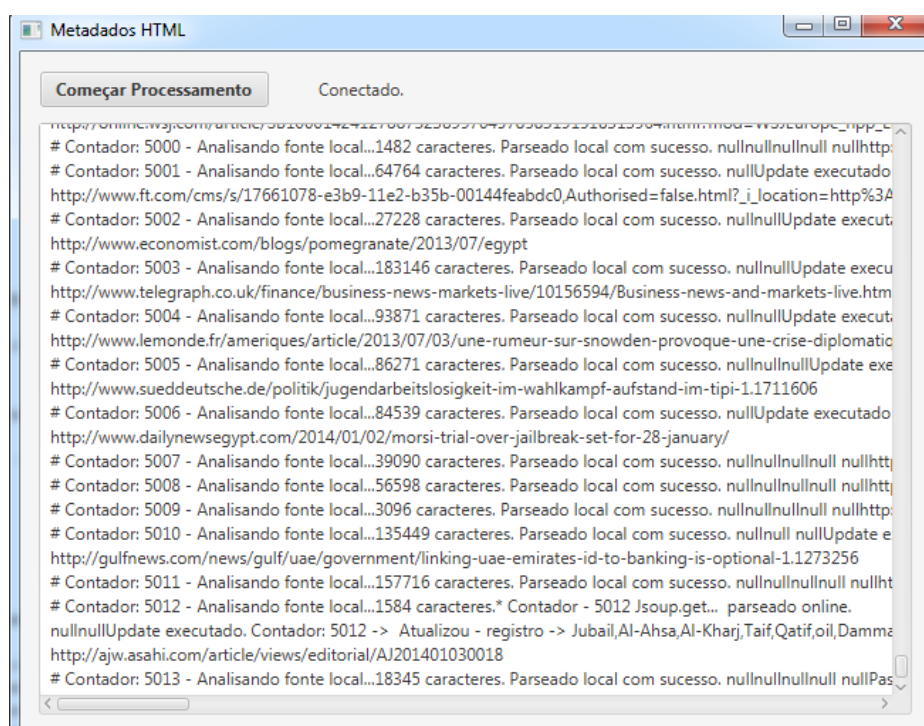


Figura 61: Interfaz de MetadadosHTML (v2)

MetadadosHTML (v2) identificó 20.801 palabras clave en el conjunto de la muestra, 3.861 únicas. Se detectaron 80 errores -datos numéricos y frases utilizadas como palabras clave- que suponen un 2,07% sobre el total, y 13 palabras clave que MetadadosHTML no identificó (0,33%). El resultado final fue 3.846 términos únicos.

48 Analizador de código HTML. Más información en <https://jsoup.org/> (último acceso: 20/05/2017)

4. RESULTADOS

Además, tanto en la primera versión de MetadadosHTML como en la segunda, los datos reflejan la variabilidad de la problemática descrita anteriormente: campos vacíos, porque no son utilizados por un periódico, y ruido informativo, por la inclusión de datos que no corresponden al metadato que se analiza. A pesar de lo anterior, MetadadosHTML se revela como herramienta útil para la extracción de metadatos del código fuente de noticias, así como para la búsqueda de noticias por palabras clave.

4.5.1.2. DEFINICIÓN DE LA SUBMUESTRA₂

Utilizando una implementación de los resultados de MetadadosHTML, se lanzaron varias consultas aleatorias relacionadas con los temas identificados en 4.1.3. DEFINICIÓN DE SUBMUESTRAS para la validación de ONAPE. A continuación, se describen brevemente cada uno de los casos, así como las noticias en que se trata.

Tema 1: DISPUTA TERRITORIAL DE LAS ISLAS SENKAKU

La disputa territorial por las Islas Senkaku/Diaoyou es un conflicto que afecta a varios Estados: Japón, China, Corea del Sur y Taiwan. Estos llevan años de tensión política por la pertenencia de este territorio a uno u otro, pues cada Estado se lo atribuye en virtud de distintos criterios. Algunas de las noticias relativas a esta *AccionConflicto* son las que se muestran en la Tabla 19.

N	Periódico	Titular	Palabras clave
1	<i>The Asahi Shimbun</i> (18/07/2013)	Los medios chinos acusan a Abe de política peligrosa - AJW por <i>The Asahi Shimbun</i>	AJW, <i>The Asahi Shimbun</i> , Japón, noticias, análisis, terremoto, Fukushima, crisis nuclear, China, Shinzo Abe, Elecciones de la Cámara Alta, Islas Senkaku
2	<i>The Asahi Shimbun</i> (30/08/2013)	Encuesta: sólo el 48% sabe de la propiedad "indiscutible" de Japón de las Senkakus - AJW por <i>The Asahi Shimbun</i>	AJW, <i>The Asahi Shimbun</i> , Japón, noticias, análisis, Terremoto, Fukushima, crisis nuclear, Oficina del Gabinete, Islas Senkaku, China, encuesta
3	<i>The Times of India</i> (27/11/2013)	China afirma su control sobre la zona aérea a pesar de los bombarderos B-52 estadounidenses	Bombas US B-52, El Pentágono, tensión, Islas Senkaku, Mar de la China Meridional, Diaoyus, China afirma su control sobre zona aérea
4	<i>The New York Times</i> (30/11/2013)	China choca por primera vez en la nueva zona aérea	Relaciones Internacionales de los Estados Unidos, Fuerzas Militares y de Defensa, Aerolíneas y Aviones, China, Japón, Islas Senkaku, Mar de la China Meridional
5	<i>The Asahi Shimbun</i> (16/12/2013)	Análisis: Con China aún en su mente, Abe finaliza su año diplomático de 2013 - AJW by <i>The Asahi Shimbun</i>	AJW, <i>The Asahi Shimbun</i> , Japón, noticias, análisis, Terremoto, Fukushima, crisis nuclear, China, Corea del Sur, ASEAN, zona de identificación de defensa aérea, Islas Senkaku

Tabla 19: Noticias sobre la disputa territorial por las Islas Senkaku

4. RESULTADOS

Se seleccionaron cinco noticias en las que apareciera el término *Islas Senkaku (Senkaku Islands)*, restringiendo el tema al conflicto por su pertenencia como territorio a un Estado u otro.

Tema 2: HOMENAJES A NELSON MANDELA TRAS SU MUERTE

Nelson Mandela muere en Johannesburgo, Sudáfrica, el 5 de diciembre de 2013. Se celebran homenajes en países de todo el mundo, destacando su papel tanto en la historia y política nacional de Sudáfrica como en el plano internacional. Algunas de las noticias relativas a esta AccionConflicto son las que se muestran en la Tabla 20.

N	Periódico	Titular	Palabras clave
6	<i>The New York Times</i> (06/12/2013)	La muerte de Mandela sacude el sentimiento por su pérdida alrededor del mundo - NYTimes.com	Apartheid (Política), Mandela Nelson, William Duque de Cambridge, Singh Manmohan, Cameron David, Tutu Desmond M, Sudáfrica
7	<i>El Univesal</i> (06/12/2013)	<i>El Univesal</i> - El Mundo - Nelson Mandela será enterrado el 15 de diciembre en Qunu	Mandela, funeral, diciembre, qunu, el mundo, internacional, noticias mundo, noticia, <i>El Univesal</i>
8	<i>Gulf News</i> (10/12/2013)	World unites for Nelson Mandela memorial <i>Gulf News.com</i>	Sudáfrica, Pretoria, Legado de Mandela, Sudáfrica
9	<i>The Daily Telegraph</i> (10/12/2013)	Servicio en memoria a Nelson Mandela: líderes mundiales y condolientes se reúnen para dar el adiós final	Nelson Mandela, Madiba, Sudáfrica, Soweto, Stadio
10	<i>Süddeutsche Zeitung</i> (10/12/2013)	Nelson Mandela ha muerto y el mundo occidental está de duelo. Sin embargo, el prestigio de Nelson Mandela no siempre fue tan alto como ahora	Nelson Mandela, Sudáfrica, Franz Josef Strauss, Ronald Reagan, Margaret Thatcher, Apartheid, Estados Unidos, <i>Süddeutsche Zeitung</i> , SZ

Tabla 20: Noticias en las que aparece Mandela como palabra clave

Se seleccionaron cinco noticias en las que apareciera el término *Mandela*, restringiendo la temática a los funerales por la muerte del expresidente de Sudáfrica.

Tema 3: RESTRICCIONES A LA ENTRADA DE INMIGRANTES EN REINO UNIDO Y ESTADOS UNIDOS

Se debaten en Reino Unido las restricciones de entrada al territorio a inmigrantes de Europa del Este. Las medidas son bien acogidas por algunos sectores, mientras que son rechazadas por otros. En Estados Unidos, en el mismo periodo, se plantea una modificación de la ley de inmigración. La Tabla 21 muestra algunas de las noticias sobre este asunto, publicadas en *The Daily Telegraph* entre 2013 y 2014. La noticia 15, de *The New York Times*, trata un asunto relacionado en Estados Unidos.

4. RESULTADOS

N	Periódico	Titular	Palabras clave
11	<i>The Daily Telegraph</i> (16/12/2013)	El límite a la migración en la Unión Europea de Theresa May es “tuerto” e ilegal, sugiere Nick Clegg - Telegraph	Primer Ministro, David Cameron, inmigrantes ilegales, inmigración, Europa del Este, inmigración, noticias de Reino Unido, noticias
12	<i>The Daily Telegraph</i> (24/12/2013)	Nigel Farage: no preste atención a la advertencia de "ríos de sangre" de Vince Cable - Telegraph	Inmigración, noticias de Reino Unido, noticias
13	<i>The Daily Telegraph</i> (30/12/2013)	Inmigrantes búlgaros reciben lecciones para aprender a explotar los beneficios británicos - Telegraph	Bulgaria, Rumanía, migrantes, UE, beneficios, explotación, lecciones, Sofía, restricciones, inmigración, noticias de Reino Unido, noticias
14	<i>The Daily Telegraph</i> (01/01/2014)	Ciudadanos no europeos podrán trabajar en Reino Unido tras el levantamiento de las restricciones a búlgaros - Telegraph	Inmigración, noticias de Reino Unido, noticias
15	<i>The New York Times</i> (02/01/2014)	Boehner cambiará de nuevo sobre la inmigración - NYTimes.com	John Boehner, Rebecca Tallent, inmigración, inmigración ilegal, Legislación, Cámara de los Representantes, Senado, ciudadanía

Tabla 21: Noticias relativas al rechazo de inmigrantes en Reino Unido

Se seleccionaron 5 noticias en las que apareciera el término *Inmigración (Immigration)*. Posteriormente, se restringió el tema a las restricciones a la inmigración en Reino Unido y Estados Unidos

Tema 4: CASO DE CORRUPCIÓN EN TURQUÍA

En Turquía, Recep Tayyip Erdogan y su gobierno se ven afectados por un escándalo de corrupción. El Primer Ministro niega la existencia de corrupción y denuncia que se trata de una conspiración internacional en la que participan distintas entidades, dentro y fuera del Estado, para derrocarlo. La Tabla 22 recoge algunas de las noticias sobre esta cuestión. El periódico más leído en el periodo de toma de la muestra, *Today's Zaman*, fue clausurado en marzo de 2016 (Akarcesme, 2016).

N	Periódico	Titular	Palabras clave
16	<i>Today's Zaman</i> (17/12/2013)	Las acciones turcas caen y la lira se debilita tras las investigaciones por corrupción	Turquía, noticias turcas, lira turca, dólar, corrupción, Erdogan
17	<i>Today's Zaman</i> (25/12/2013)	PM [Primer Ministro] amenaza a empresas, medios de comunicación y grupos civiles en medio de los problemas de corrupción	Turquía, noticias turcas, Erdogan, corrupción
18	<i>The Economist</i> (04/01/2014)	Corrupción en Turquía: El camino árabe – <i>The Economist</i>	Corrupción en Turquía, Política, Turquía, Política internacional, Política turca, Recep Tayyip Erdogan

Tabla 22: Noticias sobre caso de corrupción en Turquía

4. RESULTADOS

Se seleccionaron tres noticias en las que apareciera el término corrupción (*corruption*). Para restringir la temática, se emplean también los términos Turquía (*Turkey*). De otras noticias que trataban este asunto no fue posible recuperar todas las palabras clave de su código fuente, ni con el software MetadadosHTML ni accediendo posteriormente a su página, debido al cierre del periódico.

Tema 5: MERCADO ENERGÉTICO

En la Tabla 23 se muestran varias noticias de ámbito internacional sobre el mercado energético. Para esta selección de cinco noticias, se empleó como palabra clave *Energía (Energy)*.

N	Periódico	Titular	Palabras clave
19	<i>The Times of India</i> (11/07/2013)	El proyecto Kudankulam consigue la aprobación para la generación de energía - <i>The Times of India</i>	Reactor Voda Voda Energo, Tamil Nadu, disputa Kudankulam, Proyecto de energía nuclear Kudankulam, KUDANKULAM, Junta Regulatoria de la Energía Atómica, energía atómica
20	<i>China Daily</i> (12/08/2013)	La nueva política verde da un gran impulso a las industrias Economía chinadaily.com.cn	Innovación tecnológica, verde, energía, Red eléctrica, baja emisión de carbono
21	<i>China Daily</i> (06/11/2013)	Fábricas se vuelven verdes y eficientes Industrias chinadaily.com.cn	Energía, calidad del aire, polución, energía verde
22	<i>The Asahi Shimbun</i> (06/12/2013)	El Ministro de Economía dice que la energía nuclear permanece como una importante fuente de energía - AJW by <i>The Asahi Shimbun</i>	AJW, <i>The Asahi Shimbun</i> , Japón, noticias, análisis, terremoto, Fukushima, crisis nuclear, energía nuclear, Toshimitsu Motegi, Plan Básico de Energía, fuentes de energía renovable, METI, Ministro de Industria, DPJ, LDP
23	<i>The Daily Telegraph</i> (02/01/2014)	Ofgem dice que los mercados energéticos son demasiado opacos como para saber si los clientes están siendo estafados - Telegraph	Energía, La Tierra

Tabla 23: Noticias sobre mercado energético

El siguiente apartado describe el proceso de instanciación de la ontología ONAPE para su validación.

4.5.2. IDENTIFICACIÓN DE RELACIONES SEMÁNTICAS NOTICIA-ONAPE PARA LA VALIDACIÓN DE ONAPE

Durante su modelado, evaluación y utilización, una ontología, describe y recupera información con mejores resultados si los documentos a partir de los que se crea y aquellos sobre los que se aplica son del mismo periodo y tipo (Hlomani, 2014). Por ello, en el modelado de ONAPE se emplearon términos de indización de la submuestra₁.

4. RESULTADOS

Para evaluar la capacidad de ONAPE para ser instanciada, se identificaron las relaciones entre términos de noticias de la submuestra₂ y los elementos de ONAPE. Estas relaciones están basadas en Baños-Moreno (2013). Así, para cada objeto analizado, se construyeron colecciones de palabras clave y se instanciaron con ONAPE:

- **Relación 1 (esUn):** cuando las palabras clave son identificadores o nombres propios, que en una ontología son representados mediante instancias, que tienen su **clase** de ONAPE. También se emplean, como palabras clave, nombres comunes, que en una ontología son representados como clases.
- **Relación 2 (equivalencia exacta):** las palabras clave de las noticias y las clases de la ontología coinciden en su significado, con o distinto significantes.
- **Relación 3 (jerarquía de clase).** Puede ser de tipo **3.1 (clase-subclase)**, donde los términos de las noticias son más genéricos que las clases de ONAPE; o **3.2 (subclase-clase)**, cuando las palabras clave de la noticia son más específicas que las clases de ONAPE.
- **Relación 4 (jerarquía de parte)**, que puede ser de dos tipos: **4.1 (todo-parte):** las palabras clave son el todo y los elementos de ONAPE son la parte; **4.2 (parte-todo):** las palabras clave son la parte y los elementos de ONAPE el todo.
- **Relación 5 (asociativa):** existe algún tipo de relación semántica entre la palabras clave y algunas clases de la ontología distinta a las especificadas anteriormente.
- **Relación 6 (inexistente):** no se encuentra ninguna relación entre el término de la submuestra₂ y ONAPE.

Posteriormente, a partir de los resultados obtenidos, cada validación es analizada a partir de precisión, exhaustividad y medida-F, descritas en 2.3.4. EVALUACIÓN DE LA INDIZACIÓN. Para aplicar estas medidas, se adaptan los componentes de cada ecuación, de la siguiente forma:

- Total Documentos Relevantes Recuperados = Total (Relaciones 1 + Relaciones 2). De esta forma, se tiene en cuenta otra medida de evaluación, la precisión de la indización ya que, al tener en cuenta sólo este tipo de relaciones, se le da preferencia a los conceptos más específicos para representar el contenido de un documento y no a los más generales (Cleveland y Cleveland, 2013, p. 159).
- Total Documentos Relevantes del Sistema = Total de palabras clave
- Total Documentos Recuperados = Total Relaciones encontradas, a excepción de la de tipo 6, esto es, no se ha identificado ninguna relación entre la palabra clave y algún elemento de ONAPE

Así, las medidas de evaluación de ONAPE son:

$$\text{Exhaustividad} = \frac{\text{Total}(\text{Relaciones 1} + \text{Relaciones 2})}{\text{Total de palabras clave}}$$

Figura 62: Fórmula para el cálculo de la exhaustividad de ONAPE

$$\text{Precisión} = \frac{\text{Total}(\text{Relaciones 1} + \text{Relaciones 2})}{\text{Total Relaciones encontradas}}$$

Figura 63: Fórmula para el cálculo de la precisión de ONAPE

$$F = \frac{\text{Precisión} \times \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}}$$

Figura 64: Fórmula para el cálculo de la medida-F de ONAPE

Exhaustividad y precisión arrojan valores de 0 a 1. Por contra, la medida-F, que mide la armonía entre las medidas anteriores, debe tener un valor alrededor del 0,50, Cuanto más cerca se está de esta cifra, mayor armonía existe entre exhaustividad y precisión.

Los resultados pueden verse mejorados si, además de las relaciones 1 (instancia-clase) y 2 (clase-clase), se tienen en cuenta las de tipo 3 (jerarquía de clase) y 4 (jerarquía de parte), ya que un término más genérico (relación 3.2), así como el todo de una relación todo-parte (relación 4.2) también representan el concepto, aunque con un peso menor (ver Figura 65).

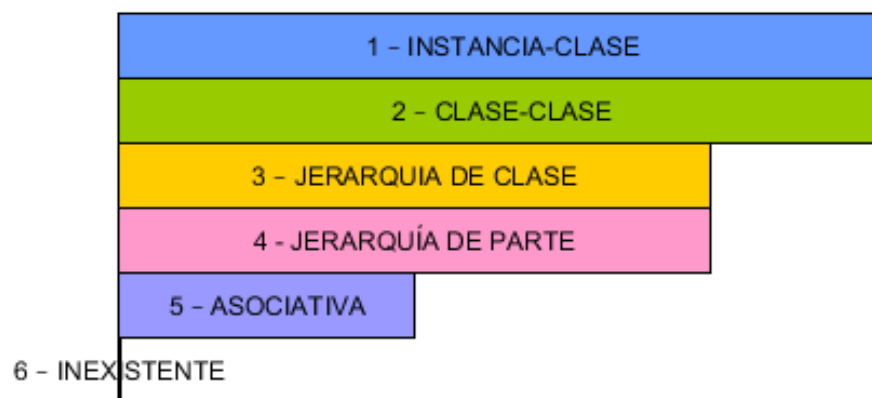


Figura 65: Tipos de relación entre palabras clave y ONAPE y grado de importancia

En este sentido, puede hablarse de una exhaustividad y precisión exactas y una exhaustividad y precisión cercanas (Baños-Moreno, 2013), que tienen en cuenta, para el cálculo, las relaciones tipo 3 y 4.

La validación de ONAPE se aplica en dos procesos: sobre las palabras clave extraídas del código fuente de noticias de la submuestra₂ y sobre el texto completo de parte de dicha

muestra. El siguiente apartado describe el proceso de instanciación de ONAPE para su validación a partir de las palabras clave extraídas del código fuente de las noticias.

4.5.3. VALIDACIÓN DE ONAPE MEDIANTE INSTANCIACIÓN

Las Tablas que se muestran a continuación representan, para cada tema, las palabras clave extraídas del metadato *Keywords* (columna 1), siendo suprimidas aquellas que tenían el mismo significante. La relación semántica que se establece entre éstas y las clases de ONAPE (columna 3) se muestran en la columna 2, siguiendo la numeración expresada en el párrafo anterior. Se destaca con un asterisco aquellos elementos para los que existe una anotación posterior.

Tema 1: DISPUTA TERRITORIAL DE LAS ISLAS SENKAKU

Palabra clave	Relación	Elementos ONAPE
AJW	1*	Sigla
Aerolíneas y Aviones	2 2	Aerolinea Avion
Análisis	1	AccionInvestigacion
<i>The Asahi Shimbun</i>	1*	EmpresaMediatica
ASEAN	1*	Sigla
Bombas US B-52	1	Bomba
China	1*	Estado
China afirma su control sobre zona aérea	1*	DisputaTerritorialDeLasIslasSenkaku
Corea del Sur	1*	Estado
Crisis nuclear	1*	CrisisNuclear
Diaoyus	1	Isla
El Pentágono	1*	FuerzasArmadas
Elecciones de la Cámara Alta	1	AccionEleccionesPoliticasNacionales
Encuesta	1	AccionEncuesta
Fuerzas Militares y de Defensa	2	FuerzasArmadas
Fukushima	1	AccidenteNuclear
Islas Senkaku	1	Isla
Japón	1*	Estado
Mar de la China Meridional	1	Mar
Noticias	1*	Género periodístico
Oficina del Gabinete	1 / 4.2*	Gobierno

Palabra clave	Relación	Elementos ONAPE
Relaciones Internacionales de los Estados Unidos	1	AccionRelacionesInternacionales
Shinzo Abe	1*	Persona
Tensión	2	Tension
Terremoto	2*	AccionTerremoto
Zona de Identificación de Defensa Aérea ⁴⁹	1	Zona de Seguridad

Tabla 24: Instanciación de la submuestra₂ en ONAPE. Tema 1

Las cinco noticias recogen, en total, 26 términos, de las cuales 21 son instancias y 5 son clases. La instancia *China afirma su control sobre zona aérea* es entendida como una acción parte de una mayor, *Disputa Territorial De Las Islas Senkaku*, un tipo de *AccionConflicto*.

Varias palabras clave son ambiguas, ya que pueden referirse a diferentes elementos que no quedan claros ni siquiera en el texto completo. Esto tiene que ver con la homonimia, es decir, un mismo significante para distintos significados. Se toma una decisión sobre el significado al que se refiere en cada caso y se explica el porqué de la misma.

Se analizan a continuación otras cuestiones relevantes:

- La sigla *AJW* se relaciona con *The Asahi Shimbun* mediante *esSiglaDe* (ver Figura 66). Otra forma de representar las siglas es con una propiedad tipo dato. Se describe la primera forma por su facilidad de ser empleada como punto de acceso

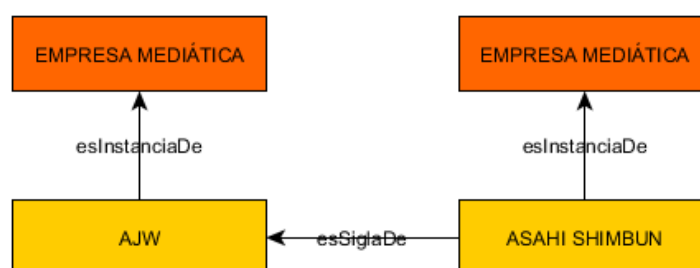


Figura 66: Relación entre las instancias *AJS* y *The Asahi Shimbun*

- De forma similar a la anterior palabra clave, la sigla *ASEAN* se relaciona con *Association of Southeast Asian Nations* mediante *esSiglaDe*. Dicha organización, por su parte, es una instancia de *OrganizacionEconomicoFinanciera*

⁴⁹ [Programa por el que] «China exige a todo avión que sobrevuele dicha zona que informe con antelación al Ministerio de Exteriores o la Administración de Aviación Civil de los planes de vuelo, que mantenga abiertas las comunicaciones de radio y responda "de una forma puntual y precisa" a las preguntas de identificación de las autoridades chinas» (Europa Press, 2013)

4. RESULTADOS

- *China, Corea del Sur y Japón* pueden ser Estados (organización política) o países (espacio físico). Como ejecutores de la acción, se entiende que, en este caso, son Estados. La dificultad para identificar esta cuestión es característica del lenguaje natural, la ambigüedad a nivel semántico (Escolano Ruiz et al., 2003, p. 206).
- *China afirma su control sobre zona aérea esParteDe* una acción mayor, la *DisputaTerritorialDeLasIslasSenkaku*.
- *Crisis nuclear*, aunque ha sido considerada como clase, hace referencia a una crisis nuclear específica. Podría ser considerada una instancia. Para ello, requeriría un significante distinto al empleado por los periódicos como palabra clave.
- La instancia *El Pentágono* se refiere a *Edificio*. Sin embargo, como agente activo en la noticia, debe ser entendida como sinónimo del *Departamento de Defensa de los Estados Unidos*⁵⁰. Esta organización es una instancia de *FuerzasArmadas*.
- La instancia *Fukushima* puede referirse a la prefectura de Fukushima, en Japón, cualquiera de las dos centrales nucleares ubicadas en dicha prefectura, Fukushima I y Fukushima II, ubicadas en la prefectura de Fukushima, al Accidente nuclear de Fukushima I, iniciado el 11 de marzo de 2011. Probablemente, el medio AJW se refiere a este último, una instancia de *AccidenteNuclear*.
- *Noticias* es uno de los géneros de los medios de comunicación. Otros que suelen aparecer son artículos de opinión, entrevista o reportaje.
- *La Oficina del Gabinete [de Japón] esParteDel* Gabinete de Japón, por tanto, puede ser entendida como parte de *Gobierno*.
- *Relaciones Internacionales de los Estados Unidos* es una acción, el acto por el que un Estado se relaciona con otro. Por esta cuestión, es una instancia de *AccionRelacionesInternacionales*.
- *Shinzo Abe* es una instancia de *Persona*. La ontología también recoge la clase *JefeDeGobierno*, subclase de *RolPolitico*. *Shinzo Abe* podría relacionarse, mediante *tieneRol*, con *JefeDeGobierno*, directamente o mediante otra instancia, como *JefeDeGobiernoJapones*, por ejemplo. Sería conveniente especificar el periodo en que se mantiene dicho rol mediante propiedades tipo *fechaDeInicio* y *fechaDeFin*. Estas decisiones debe tomarlas quien use la ontología.
- *Terremoto* ha sido considerada clase, aunque, al referirse a un terremoto concreto, podría ser clasificado como instancia. Requiere un significante distintivo.

50 «Departamento del brazo ejecutivo del gobierno federal de Estados Unidos encargado de coordinar y supervisar todas las agencias y funciones del gobierno relacionadas directamente con la seguridad nacional y las Fuerzas Armadas de los Estados Unidos» (Wikipedia, 2017)

Tema 2: HOMENAJES A NELSON MANDELA TRAS SU MUERTE

Palabra clave	Relación	Clase ONAPE
Apartheid	2*	Sistema
Apartheid (Política)	2*	Sistema
Cameron David	1*	Persona
Diciembre	1	Mes
El Mundo	1	Sección [de periódico]
<i>El Univesal</i>	1	EmpresaMediatica
Estadio	2	Edificio
Estados Unidos	1*	Estado
Franz Josef Strauss	1*	Persona
Funeral	4.2*	AccionEventoSocial
Internacional	1	Sección [de periódico]
Legado de Mandela	6*	Herencia
Madiba	1*	Persona
Mandela	1*	Persona
Mandela Nelson	1*	Persona
Margaret Thatcher	1*	Persona
Nelson Mandela	1*	Persona
Noticia	1	Género periodístico
Noticias Mundo	1	Sección [de periódico]
Pretoria	1	CiudadCapital
Qunu	1	Ciudad
Ronald Reagan	1*	Persona
Singh Manmohan	1*	Persona
Soweto	1	Distrito
Sudáfrica	1*	Estado
Süddeutsche Zeitung	1	EmpresaMediatica
SZ	1	Sigla
Tutu Desmond M	1*	Persona
William, Duque de Cambridge	1*	Persona

Tabla 25: Instanciación de la submuestra₂ en ONAPE. Tema 2

Se distinguen los siguientes aspectos en relación con las palabras clave de la Tabla 25:

- *Apartheid* y *Apartheid (Política)* hacen referencia a un mismo elemento, un sistema de segregación racial, *especialmente el establecido en la República de Sudáfrica por la minoría blanca* (RAE, 2014a). Se trata, pues, de un tipo de *Sistema*. Sin embargo, teniendo en cuenta la definición, podría incorporarse una nueva clase, subclase de *Sistema*: *SistemaDeSegregacion*

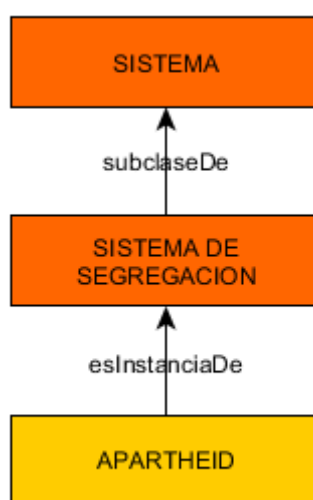


Figura 67: Instanciación de ONAPE con la instancia Apartheid

- *Cameron David, Franz Josef Strauss, Madiba, Mandela, Mandela Nelson, Margaret Thatcher, Nelson Mandela, Ronald Reagan, Singh Manmohan, Tutu Desmond M y William, Duque de Cambridge* son *Personas* que participan en la acción *Funeral* [en honor a Nelson Mandela] por su *RolPolitico*, o bien en otras acciones descritas en las noticias de la submuestra₂. Por ejemplo: a) *Cameron David tieneRol JefeDeGobierno* [de Reino Unido], que mantiene entre 2010 y 2016; b) *Margaret Thatcher tieneRol JefeDeGobierno* [de Reino Unido] entre 1979 y 1990; c) *Nelson Mandela tieneRol Activista* [desde 1943] y *tieneRol JefeDeGobierno* [de Sudáfrica] entre 1994 y 1999. Esta persona es representada de diversas formas: *Madiba*, sobrenombre con el que es/era conocido Nelson Mandela; *Mandela*; *Nelson Mandela* y *Mandela Nelson*.

Estos ejemplos revelan la importancia de presentar uniformemente los nombres propios, para facilitar tanto su descripción como su recuperación. Probablemente, esta cuestión esté resuelta en los medios de comunicación de la submuestra₂ para

4. RESULTADOS

este tema y se deba a su diversidad, cada uno con sus normas de descripción. También destaca la necesidad de incorporar nuevos componentes a ONAPE: la clase *Sobrenombre* y las relaciones *tieneSobrenombre* y *esSobreNombreDe*, o bien una propiedad tipo dato, *sobrenombre*. La elección de unos elementos u otros depende de quien use la ontología para describir sus documentos

- *Funeral* es subclase de *EventoSocial*, ha sido clasificada como clase, aunque hace referencia a una instancia del mundo real, el funeral de Nelson Mandela. Requeriría un significativo como *FuneralDeMandela*
- *Legado de Mandela* se ha clasificado como instancia de *Herencia*, concepto con dos significados relevantes para una posible incorporación a ONAPE. Ambos son recogidos por la RAE (2014f) en sus acepciones 2 y 4: a) «rasgo o rasgos morales, científicos, ideológicos, etc., que, habiendo caracterizado a alguien, continúan advirtiéndose en sus descendientes o continuadores; b) conjunto de bienes, derechos y obligaciones que, al morir alguien, son transmisibles a sus herederos o a sus legatarios». La primera acepción es aquella con la que existiría una relación de tipo 1 con *legado de Mandela*
- *Sudáfrica*: en el contexto analizado este término puede referirse bien al *Estado* de la República de Sudáfrica o bien a la población sudafricana en su conjunto, que en la ontología de dominio podría ser una instancia de la clase *Ciudadania*.

Tema 3: RESTRICCIONES DE ENTRADA DE INMIGRANTES EN REINO UNIDO

Palabra clave	Relación	Clase ONAPE
Beneficios	2	Beneficio
Bulgaria	1*	País
Cámara de los Representantes	4.2	Parlamento
Ciudadanía	2	Ciudadania
David Cameron	1	Persona
Europa del Este	1	Region
Explotación	4.2*	AccionDelito
Inmigración	4.2*	AccionMigrante
Inmigración ilegal	4.2*	AccionMigrante + Inmigrante (Rol) + Ilegalidad (Situacion)
Inmigrantes ilegales	2*	Inmigrante (Rol) + Ilegalidad (Situacion)
John Boehner	1	Persona
Lecciones	2*	AccionEducacion
Legislación	2	Ley
Migrantes	2	Migrante
Noticias	1	Sección [de periódico]
Noticias de Reino Unido	1	Sección [de periódico]
Primer Ministro	1	JefeDeGobierno
Rebecca Tallent	1	Persona
Restricciones	2	AccionRestriccion
Rumanía	1*	País
Senado	4.2	Parlamento
Sofía	1	CiudadCapital
UE	1*	Sigla

Tabla 26: Instanciación de la submuestra₂ en ONAPE. Tema 3

4. RESULTADOS

Se muestran 22 palabras clave. En cuanto a las particularidades de algunas de las palabras clave de la Tabla 26, destacan:

- *Bulgaria y Rumanía*: como antes, podrían ser estados (conjuntos de instituciones) o países (territorios). Por contexto, parece hacerse referencia al segundo caso. Se trata de dos países de los que emigran personas hacia Reino Unido.
- *David Cameron* es una *Persona* con el *RolPolitico JefeDeGobierno* [de Reino Unido], que también aparece con un término alternativo, *Primer Ministro*.
- *Explotación*: la RAE (2014d) recoge varias acepciones para *Explotar*. En el contexto del tema seleccionado, la entrada relacionada es «sacar utilidad de un negocio o industria en provecho propio».
- *Inmigración* (e *Inmigración ilegal*) es el desplazamiento de personas de un territorio a otro desde el punto de vista del territorio receptor. Cuando ese movimiento migratorio se produce en una situación de ilegalidad, esto es, contraria a ley, se habla de inmigración ilegal. ONAPE recoge la clase *AccionMigrante*, que puede concebirse como la clase de la que pende *Inmigración*.

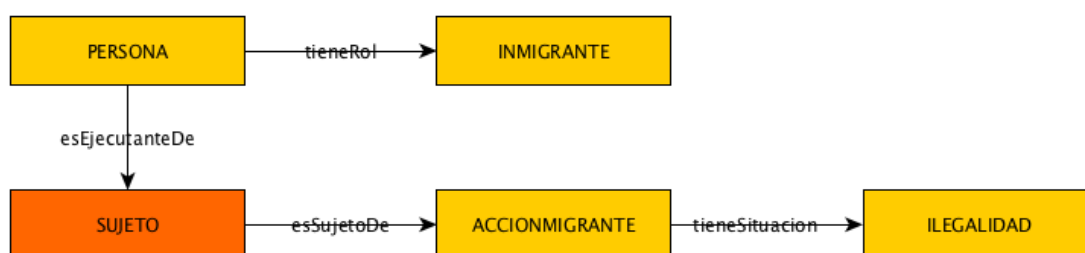


Figura 68: Acción migrante ejecutada por inmigrantes ilegales

Dependiendo de quien describe un evento, se puede presentar de distintas formas. En la ontología, inmigrante o emigrante, es decir, si las personas vienen o van, se define a partir del agente ejecutor de la acción. Por tanto, una acción de inmigración ilegal se podría definir a partir de las clases y relaciones (ver Figura 68).

- *Inmigrantes ilegales*: como se indica en el punto anterior, ONAPE contempla la situación *Ilegalidad*, que puede definirse a nivel de los *Agentes* o de las *Acciones*. Esta decisión queda a criterio de quien utilice la ontología.
- *Lecciones*: una de las noticias de la muestra destaca que *búlgaros reciben lecciones para aprender a explotar los beneficios británicos*. Se hace referencia, por tanto, a una *AccionEducacion*, aunque no se trate de una enseñanza reglada.
- *UE*: es la sigla de la instancia *UnionEuropea*. Por tanto, tendríamos dos instancias, *UE* y *UnionEuropea*, vinculadas mediante la relación *esSiglaDe* (y su inversa).

Tema 4: CASO DE CORRUPCIÓN EN TURQUÍA

Palabra clave	Relación	Clase ONAPE
Corrupción	1	AccionCorrupcion
Corrupción en Turquía	1*	AccionCorrupcion, Sección [de periódico]
Dólar	1	Moneda
Lira turca	1	Moneda
Noticias turcas	1	Sección [de periódico]
Política	1	Sección [de periódico]
Política internacional	1	Sección [de periódico]
Política turca	1	Sección [de periódico]
Recep Tayyip Erdogan / Erdogan	1*	Persona
Turquía	1	Estado

Tabla 27: Instanciación de la submuestra₂ en ONAPE. Tema 4

En total, aparecen 10 palabras clave, siendo todas instancias de clases de ONAPE. Algunos aspectos que cabe mencionar son:

- *Corrupción en Turquía* puede ser instancia de dos clases, dependiendo de cómo se considere: una acción o una sección en un periódico.
- *Dólar* y *Lira turca* son instancias de dos monedas, que pueden relacionarse con los Estados a los que pertenece mediante *esFabricadoPor EstadosUnidosDeAmerica* y *RepublicaDeTurquia*, respectivamente (ver Figura 69).

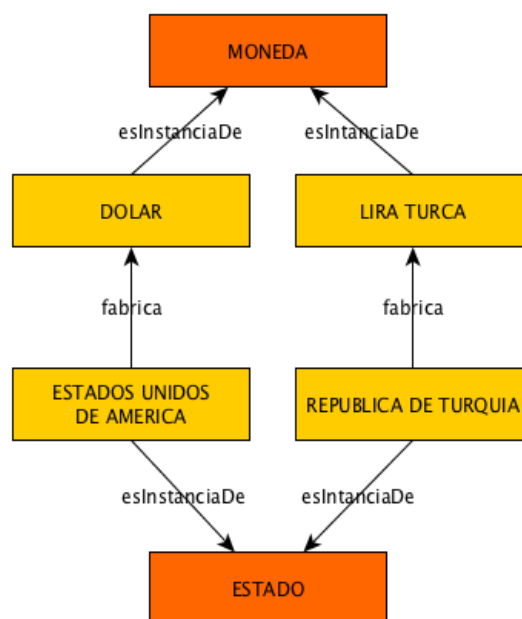


Figura 69: Relación entre Monedas y Estados para el tema Caso de Corrupción en Turquía

4. RESULTADOS

- *Recept Tayyip Erdogan (y Erdogan)* es una *Persona* con el *RolPolitico JefeDeGobierno* [de Turquía], que ostenta durante un periodo a definir entre las propiedades *fechaDeInicio* y *fechaDeFin*.

Tema 5: MERCADO ENERGÉTICO

Palabra clave	Relación	Clase ONAPE
AJW	1	Sigla
Análisis	4.2*	AccionInvestigacion
<i>The Asahi Shimbun</i>	1	EmpresaMediatica
Junta Regulatoria de la Energía Atómica	1*	OrganizacionPublica
Baja emisión de carbono	1*	AccionEmisionDeDioxidoDeCarbono + cantidad
Calidad del aire	6*	caracteristica + Calidad
Crisis nuclear	1*	CrisisNuclear
Disputa Kudankulam	1*	AccionConflicto
DPJ	1*	Sigla
Energía	2	Energia
Energía atómica	3.2	Energia
Energía nuclear	2	EnergiaNuclear
Energía verde	2	EnergiaRenovable
Fuentes de energía renovable	4.2*	RecursoEnergetico
Fukushima	1*	PlantaNuclear
Innovación tecnológica	5*	ProductoElectronico
Japón	1	Estado
KUDANKULAM	1*	PlantaNuclear
La Tierra	2	Planeta Tierra
LDP	1*	Sigla
METI	1*	Sigla
Ministro de Industria	1	Ministro
Noticias	1	Género periodístico
Plan Básico de Energía	1*	Programa
Polución	2*	Contaminante
Proyecto de planta nuclear Kudankulam	1*	PlantaNuclear + Borrador (Situacion)

4. RESULTADOS

Reactor Voda Voda Energo	1	ReactorNuclear
Red eléctrica	5*	EnergiaElectrica
Tamil Nadu	1	Persona
Terremoto	2	AccionTerremoto
Toshimitsu Motegi	1	Persona
Verde	2	EnergiaRenovable

Tabla 28: Instanciación de la submuestra₂ en ONAPE. Tema 5

Respecto de las palabras clave indicadas en la Tabla anterior, se destacan los siguientes aspectos:

- El término *Análisis*, como en diversos casos recogidos en esta validación, hace referencia a un estudio concreto, por lo que se trataría, más bien, de una instancia de una clase más específica que *AccionInvestigacion*. Para ello, sin embargo, sería necesario aportar mayor especificidad al término.
- Para representar adecuadamente el término *Calidad del aire* en ONAPE es necesario incorporar *Aire* como clase o instancia, dependiendo de cómo lo contemple quien la utilice. De esta forma, sería posible establecer una relación entre *Aire* y una subclase de *Calidad*, pendiente de definir.

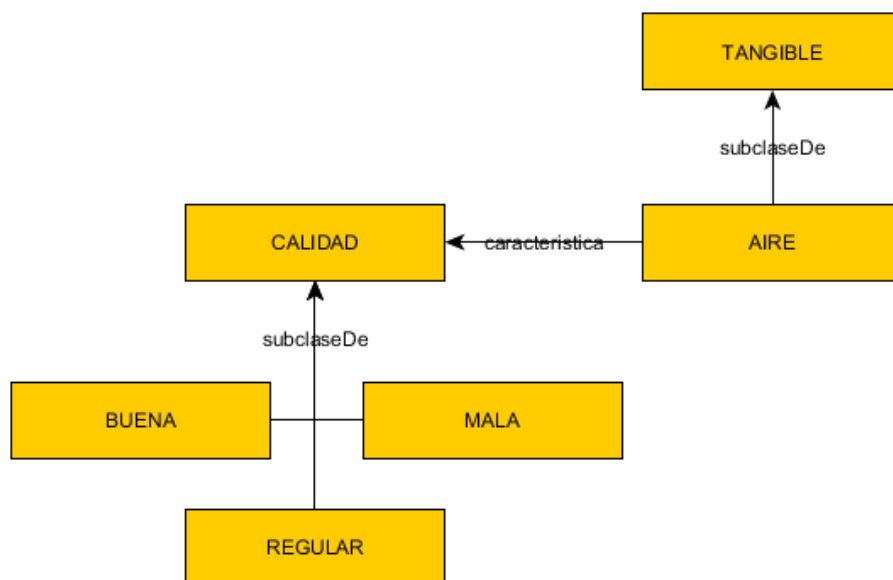


Figura 70: Relación entre Aire y Calidad y otros elementos

4. RESULTADOS

- *Junta Regulatoria de la Energía Atómica*, conocida por sus términos en inglés, *Atomic Energy Regulatory Board*, es una *OrganizacionPublica* dependiente del Gobierno de la República de la India. Al tratarse de una organización, existen otras clases que permitirían especificar algunas de sus características, tales como su alcance: *AlcanceNacional* y su ámbito de actuación: *AmbitoPublico*, entre otros.
- *Crisis nuclear*, aunque ha sido considerada como clase, hace referencia a una crisis nuclear específica. Por tanto, podría ser concebida como una instancia. Para ello, requeriría un significante distinto al empleado por los periódicos como palabra clave.
- *DPJ* y *LDP* son instancias de *PartidoPolitico* del Estado de Japón y de la República de la India: *Democratic Party of Japan* y *Liberal Democratic Party*, respectivamente.
- *Fuentes de energía renovable* es una subclase de *RecursoEnergetico*. En ONAPE se establece una relación directa entre cada recurso energético y la clase *RecursoEnergetico*. No obstante, puede ser necesaria una clase intermedia (*RecursoEnergeticoRenovable*) para referirse globalmente al conjunto de fuentes naturales de las que se puede extraer energía renovable. Queda a criterio de quien use la ontología.
- La instancia *Fukushima* aparece por el accidente nuclear de Fukushima I, iniciado el 11 de marzo de 2011. La relación entre esta central y la de *Kudankulam* tiene que ver, probablemente, con el miedo a otro accidente nuclear en una región de la India.
- *Innovación tecnológica* requeriría de mayor especificidad para identificar el elemento más adecuado en ONAPE. Se plantea la clase *ProductoElectronico* como término más aproximado.
- La instancia *KUDANKULAM* alude, probablemente, a la Central nuclear de Kudankulam, conocida por sus siglas en inglés, *Koodankulam NPP* o *KKNPP* (*Kudankulam Nuclear Power Plant*). Se trata concretamente de la «planta de energía nuclear en Koodankulam en el distrito de Tirunelveli del Estado de Tamil Nadu en el sur del país asiático de la India» (Wikipedia, 2016).
- *METI* es la sigla de *Ministry of Economy, Trade and Industry* [Repúb. de la India].
- *Plan Básico de Energía* es un *Programa* específico de energía de la República de la India que se ejecutaría a través de una *AccionPlanificacion* (ver Figura 71).



Figura 71: Relación de Plan Básico de Energía con elementos de ONAPE

- *Polución* se define como «contaminación intensa y dañina del agua o del aire, producida por los residuos de procesos industriales o biológicos» (RAE, 2014i). De acuerdo esta definición, *Polución* es sinónimo de *Contaminante*, término recogido en ONAPE. La *polución esResultadoDe* una *AccionEmisionDeContaminantes*.
- *Proyecto de planta nuclear Kudankulam* alude a la planta nuclear *KUDANKULAM*, y establece una serie de acciones para poner en funcionamiento dicha central.
- Para *Red eléctrica*, definida como «red interconectada que tiene el propósito de suministrar electricidad desde los proveedores hasta los consumidores» (Wikipedia, 2015), no existe un término equivalente en ONAPE. Sin embargo, sí existe una relación asociativa con *EnergiaElectrica*. Convendría evaluar la necesidad de incorporar el término a ONAPE o bien utilizar *EnergiaElectrica* como cuasi-sinónimo. Para ello, sería necesario contextualizar el término.

En el siguiente subapartado se recogen datos estadísticos relacionados con las palabras clave de las noticias agrupadas en los cinco temas anteriores. Asimismo, se muestra información sobre la precisión y exhaustividad de ONAPE.

4.5.3.1. DATOS ESTADÍSTICOS DEL PROCESO

La Tabla 29 recoge el total de relaciones identificadas entre palabras clave de la submuestra₂ y elementos de ONAPE, clasificadas por temas.

	Tema 1	Tema 2	Tema 3	Tema 4	Tema 5	TOTAL
Relación 1	22	24	11	10	19	86
Relación 2	5	3	7	0	7	22
Relación 3	0	0	0	0	1	1
Relación 4	0	1	5	0	2	8
Relación 5	0	0	0	0	2	2
Relación 6	0	1	0	0	1	2
TOTAL	27	29	23	10	32	121

Tabla 29: Total de relaciones entre la submuestra₂ y elementos de ONAPE por tema

4. RESULTADOS

En total se recogieron 121 palabras clave diferentes para los cinco temas analizados. De estas, más del **71%** son instancias para las que ONAPE contempla una clase específica y un **18,2%** son o equivalen a clases recogidas por la ontología. Esto significa que un **89%** de las palabras clave que aparecen en el código fuente de las noticias de la submuestra₂ pueden ser representadas por elementos de la ontología. Por otro lado, más de un **9%** de palabras clave guardan algún otro tipo de relación con elementos de ONAPE, sean relaciones de clase-subclase, todo-parte o asociativas. Es decir, para más del **98%** de las palabras clave, se halló algún tipo de relación. La excepción son los términos *Calidad del aire* y *Legado de Mandela*, que representan algo más del **1,6%** sobre el total de términos.

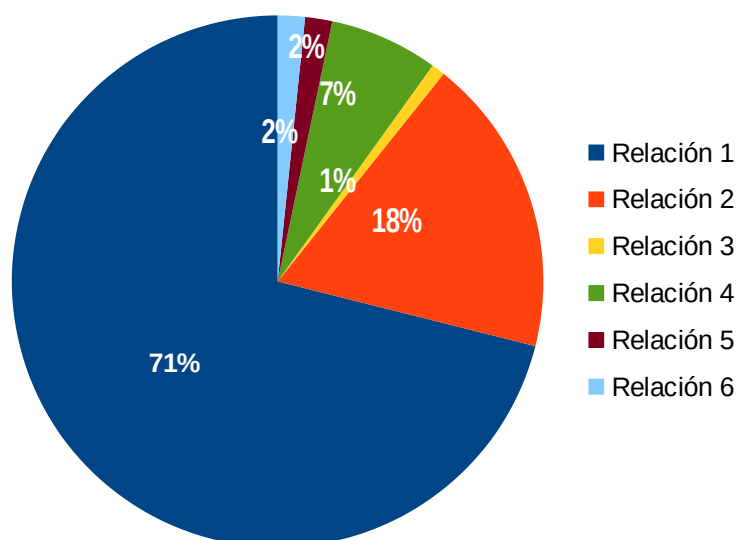


Figura 72: Medias de tipos de relaciones entre palabras clave y elementos de ONAPE

Para cada tema se calculan exhaustividad, precisión y medida-F exactas (ver Tabla 30).

	Tema 1	Tema 2	Tema 3	Tema 4	Tema 5	MEDIA
Exhaustividad (exacta)	1	0,93	0,78	1	0,81	0,90
Precisión (exacta)	1	0,96	0,78	1	0,84	0,91
Medida-F (exacta)	0,50	0,47	0,39	0,50	0,41	0,45

Tabla 30: Exhaustividad, Precisión y medida-F por temas analizados

Estos datos muestran tasas elevadas de exhaustividad, precisión y armonía entre las mismas (medida-F). En todos los casos se obtuvo un resultado superior al **78%** de exhaustividad y precisión, siendo del **100%** para los temas 1 (disputa territorial por las islas Senkaku) y 4 (caso de corrupción en Turquía).

4. RESULTADOS

En cuanto a exhaustividad, precisión y medida-F cercanas para ONAPE (ver Tabla 31), teniendo en cuenta las relaciones 1 (instancia-clase), 2 (clase-clase), 3.2 (clase-subclase) y 4.2 (todo-parte), los datos mejoran sustancialmente, especialmente para el tema 3. Exhaustividad y precisión pasan de 0,78 al 1. Esto implica el incremento de las medias globales a 0,96 y 0,98, con una mejora de la armonía entre las medidas, de 0,45 a 0,48.

	Tema 1	Tema 2	Tema 3	Tema 4	Tema 5	MEDIA
Exhaustividad (cercana)	1	0,93	1	1	0,91	0,96
Precisión (cercana)	1	0,96	1	1	0,94	0,98
Medida-F (cercana)	0,5	0,47	0,50	0,50	0,46	0,49

Tabla 31: Exhaustividad, Precisión y medida-F por temas analizados

El siguiente apartado describe el proceso de instanciación de ONAPE para su validación a partir de términos extraídos del texto completo de las noticias.

4.5.4. VALIDACIÓN POR ANOTACIÓN SEMÁNTICA

Se describe el proceso de validación mediante anotación semántica de ONA, tomada a partir del conjunto de noticias de la submuestra₂ para el tema 1, DISPUTA TERRITORIAL DE LAS ISLAS SENKAKU.

En primer lugar, se identifica cada entidad básica considerada relevante. Cada entidad es una clase de ONA (y, por tanto, de ONAPE). Para facilitar la distinción de estos elementos, se emplean los colores de marcado de la Tabla 32. Las noticias marcadas empleando estos colores se muestran en el anexo C. MARCADO SEMÁNTICO DE NOTICIAS DEL TEMA 1.

Color de marcado	Descripción
Verde	Acciones
Rosado	Agentes
Amarillo	Cosas
Púrpura	Espacios físicos
Cian	Espacios temporales
Gris	Propiedades

Tabla 32: Colores de marcado semántico para la identificación de entidades básicas en noticias

A continuación se muestran los aspectos relevantes de las noticias del tema 1 de la submuestra₂ ya marcadas semánticamente y en el idioma original (inglés). El titular se ha traducido al español. También se indica el medio que las publica, la fecha de publicación y autoría.

Aspectos semánticos relevantes de la noticia 1: China media accuses Abe of dangerous politics (Los medios chinos acusan a Abe de política peligrosa)

La noticia recoge el conjunto de acciones realizadas por unos pocos agentes. La acción punto de partida de la noticia es la acusación de *política peligrosa* por parte de medios de comunicación chinos a Shinzo Abe. En general, son citas directas e indirectas de declaraciones de Abe, incluyendo sus intenciones de cara al futuro. Esta acusación tiene su origen en estas acciones: la advertencia de Abe al Estado chino por aumentar la explotación de gas en el territorio en conflicto: el mar de la China Meridional, donde se ubican las Islas Senkaku, la visita de Abe a Ishigaki, cerca del territorio en conflicto, con el propósito de ganar las elecciones (según los medios), su intención de modificar la Constitución japonesa. Todas las acciones, en cualquier caso, forman parte de una mayor, la disputa territorial por las islas Senkaku/Diaoyou o, al menos, mantienen un estrecho vínculo con ésta.

Destaca la dificultad para simplificar estas acciones para su representación en términos de indización, ya que el texto refleja opiniones e intenciones, con alusiones a lo que alguien opina y/o quiere llevar a cabo.

Aspectos semánticos relevantes de la noticia 2: Only 48% aware of Japan's 'undisputed' ownership of Senkakus (Encuesta: sólo el 48% sabe de la propiedad "indiscutible" de Japón de las Senkakus).

Esta noticia recoge los resultados de una encuesta. Refleja, esencialmente, la opinión de la muestra consultada sobre el conflicto territorial por las islas Senkakus/Diaoyou. Es por esta cuestión por la que aparecen menos acciones y más datos numéricos, que se muestran como propiedades tipo dato. Los agentes son, principalmente, la muestra entrevista en la encuesta y los distintos subtipos de la misma, quiénes opinaban una cosa y quiénes otra. Al no analizar con tanto detalle, ONAPE no recoge varios de estos elementos, que pueden ser considerados como subclases de clases de ONAPE o como términos inexistentes.

Aspectos semánticos relevantes de la noticia 3: China asserts control over air zone despite US B-52 bombers (China afirma su control sobre la zona aérea a pesar de los bombarderos B-52 estadounidenses)

Este texto incide en tres acciones: el establecimiento de una zona de seguridad aérea por el Estado chino, cerca del territorio en conflicto, las islas Senkaku/Diaoyous, la negativa de los Estados Unidos (y Japón) a acatar las condiciones del reglamento que conlleva la zona de seguridad y sus consecuencias. Podría pensarse que estas acciones no están relacionadas con el conflicto. Sin embargo, que se establezca cerca de las islas disputadas, que los protagonistas sean precisamente China y Japón, junto con Estados Unidos, y la selección de palabras clave por el medio de comunicación (ver Tabla 19) indican lo contrario. Por

otro lado, ONAPE recoge términos generales para las acciones de esta noticia, como *AccionPrevencion*, para el establecimiento de una zona de seguridad y la vigilancia de un territorio. Por tanto, a efectos estadísticos, estas dos acciones podrían considerarse subclases de ONAPE y/o términos inexistentes.

Aspectos semánticos relevantes de la noticia 4: China scrambles jets for first time in new air zone (China despliega por primera vez en la nueva zona aérea)

Esta noticia se relaciona con las anteriores en dos aspectos: por un lado, tiene como marco el conflicto territorial por las islas Senkakus/Diaoyous; por otro, el establecimiento de una zona de seguridad aérea por el Estado chino. El texto recoge estas cuestiones, incidentes pasados y actuales con aviones y las declaraciones y críticas generadas en otros Estados por la fijación de esta zona, tensando las relaciones de China con estos. Aparecen varios modelos de aviones, que pueden recogerse como subclase de Avion, como instancias o como propiedades tipo dato, al definirlos como características de los aviones concretos que participan en la noticia. En este caso, se opta por recogerlos como instancias. Esto, junto con las distintas acciones de prevención recogidas en el texto: monitorización del espacio aéreo, identificación de objetos, patrullas aéreas, notificación de entrada en el espacio, etc. reflejan la granularidad que puede alcanzar cualquier vocabulario de representación de contenidos.

Aspectos semánticos relevantes de la noticia 5: ANALYSIS: With China still on his mind, Abe ends diplomatic work for 2013 (Análisis: Abe finaliza su año diplomático de 2013 con China aún en mente)

Este texto es el que recoge la mayor variedad de acciones y es también es el texto de mayor tamaño. Se centra principalmente en las visitas oficiales de Shinzo Abe a otros países y en diversos aspectos económicos de Japón. Entre ellos, la construcción de reactores nucleares por empresas japonesas en el extranjero que, con el recuerdo del accidente nuclear de Fukushima, genera rechazo entre la población. También destaca la cumbre de la ASEAN (*Association of Southeast Asian Nations*). En ésta, se realizan diversas declaraciones por los líderes de los estados que forman parte de la organización. Shinzo Abe rechaza el establecimiento de la zona de seguridad aérea de China en el mar de la China Meridional, vinculando una vez más la noticia con la disputa territorial por las islas Senkakus/Diaoyous.

La dificultad de esta pieza reside en la cantidad de acciones a representar, así como en establecer relaciones entre las mismas. Esta es precisamente una de las ventajas que aportan las ontologías frente a otros tipos de vocabularios. También aparecen diversas propiedades tipo dato que hay que vincular con clases.

4. RESULTADOS

En el siguiente apartado se muestran los datos estadísticos de este proceso de validación, empleando como referencia las relaciones semánticas entre texto de la noticia y ONAPE indicadas en 4.5.2. IDENTIFICACIÓN DE RELACIONES SEMÁNTICAS NOTICIA-ONAPE PARA LA VALIDACIÓN DE ONAPE.

4.5.4.1. DATOS ESTADÍSTICOS DEL PROCESO

La Tabla 33 recoge los resultados de la anotación semántica del texto de las noticias del tema 1: DISPUTA TERRITORIAL DE LAS ISLAS SENKAKU con ONA y ONAPE. Todas las acciones que aparecen en las noticias de este tema forman parte de una acción mayor, la propia disputa, o bien son colaterales a la misma. Por ejemplo, el establecimiento de una zona de seguridad aérea por parte del gobierno chino. Por esta cuestión, no se han tenido en cuenta las relaciones de jerarquía todo-parte (4).

La primera columna muestra las principales clases de la ontología ONA y ONAPE, mientras que la segunda hace referencia al total de clases distintas de ONAPE que fueron identificadas en los textos. Para cada una de estas clases también se indica la frecuencia total de clases únicas, el número de instancias distintas y su frecuencia. Una clase o instancia distinta puede tener una frecuencia de aparición en las noticias mayor a 1. Por ejemplo, *AccionDeclaracion* es un tipo de acción que será contabilizada una vez en la columna 2. Ahora bien, una misma noticia puede referirse a una misma *AccionDeclaracion* 8 veces. En este caso, se contabilizará, además, una vez en la columna 3 y 8 veces en la columna 4. Si esa noticia recoge otra *AccionDeclaracion*, realizada por una persona distinta, los contadores se incrementarán en 0 en la columna 2, en 1 en la columna 3 y tantas veces como se haga referencia a esa declaración en la columna 4.

ONA	F. CLASES ÚNICAS ONAPE	F. INSTANCIAS ÚNICAS	F. TOTAL INSTANCIAS
Acciones	44	103	200
Agentes	15	75	199
Cosas	41	72	112
Espacios físicos	7	31	99
Espacios temporales	7	25	35
TOTAL CLASES	114	306	645
Propiedades	12	64	80
TOTAL PROPIEDADES	126	370	725

Tabla 33: Total de relaciones entre elementos del tema 1 de la submuestra₂ y ONAPE

Según estos datos, se han identificado más acciones (44), seguidas de cosas (41) y agentes (15). Acciones y cosas representan el 67,4% del total de elementos. Las entidades menos frecuentes son espacios físicos y temporales (7).

4. RESULTADOS

Además, la mayoría de elementos son instancias. De nuevo, en el caso de instancias (únicas), se identificaron más elementos de tipo acción (103), seguidos de cerca por agentes (75) y cosas (72). Entre los tres representan más de un 67% del conjunto de instancias únicas. Sólo las acciones suponen casi un 27% del total de elementos.

Por último, en cuanto a la frecuencia total de instancias, prácticamente se identificaron la misma cantidad de acciones (200) y agentes (199), lo que supone una tasa media de repetición de 1,94 y 2,65, respectivamente. Este dato destaca especialmente en el caso de los espacios físicos que, con 31 instancias únicas, multiplica por 3 la tasa media de repetición (3,19). Esto es debido a las continuas alusiones a los territorios en conflicto: islas Senkakus o Diaoyous y mar de la China Meridional, entre otros.

Por lo que respecta a las propiedades, tanto de dato como de objeto, se identificaron 12 tipos, con 64 instancias únicas y 80 en total, lo que indica una baja frecuencia. Estos datos son representados gráficamente en la Figura 73. Probablemente en un análisis detallado se identificarían más propiedades.

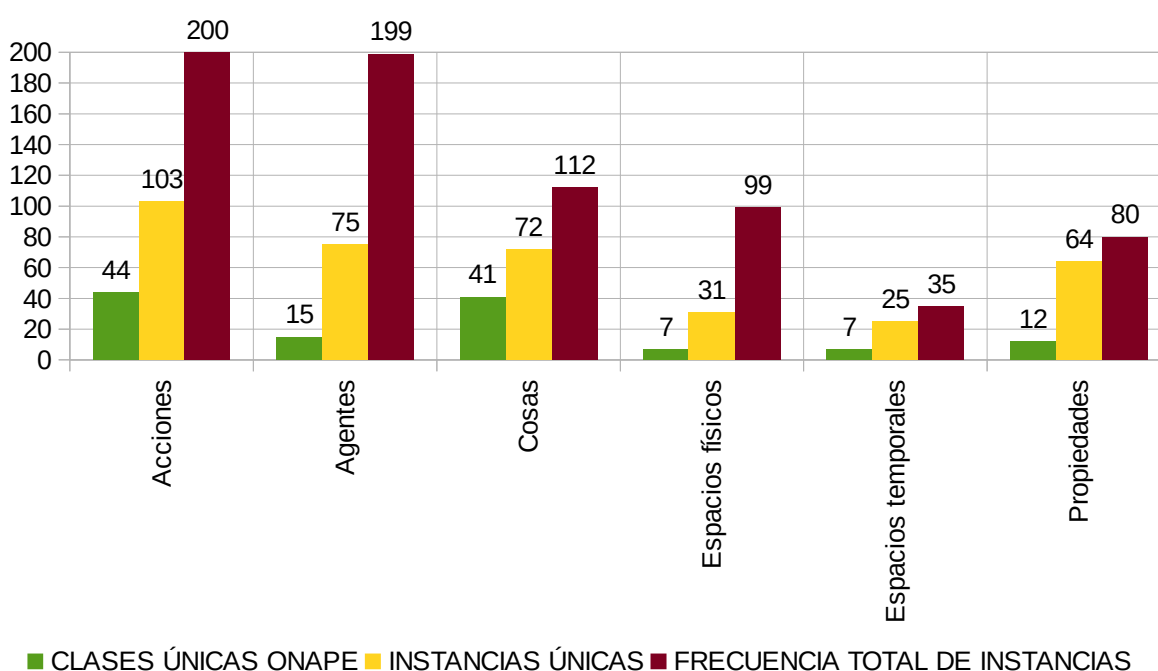


Figura 73: Clases e instancias identificadas en las noticias del tema 1 de la submuestra₂

Por otro lado, la Tabla 34 muestra el número de relaciones identificadas entre clases e instancias únicas de las noticias del tema 1 de la submuestra₂ y ONAPE por noticia. Se han tenido en cuenta 323 elementos diferentes para las cinco noticias.

4. RESULTADOS

	Noticia 1	Noticia 2	Noticia 3	Noticia 4	Noticia 5	TOTAL
Relación 1	3	1	1	1	4	10
Relación 2	61	28	44	56	117	306
Relación 3	0	0	0	0	0	0
Relación 4	0	0	0	0	0	0
Relación 5	0	0	0	0	0	0
Relación 6	0	4	1	2	0	7
TOTAL	64	33	46	59	121	323

Tabla 34: Relaciones entre términos de noticias y elementos de ONAPE por noticia

De acuerdo con la Tabla 34, más del 94% son instancias para las que ONAPE contempla una clase más o menos específica y un 3% son o equivalen a clases de la ontología. Esto significa que un 97% de las palabras clave que aparecen en el código fuente de las noticias de la submuestra₂ pueden ser representadas por elementos de la ontología.

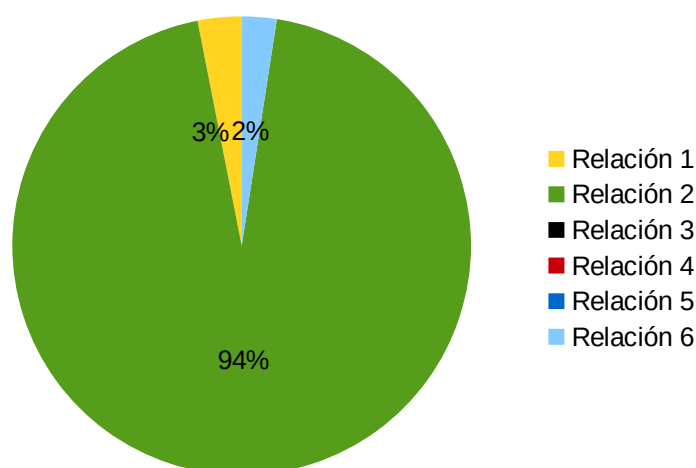


Figura 74: Medias de tipos de relaciones entre elementos de noticias y ONAPE

La excepción la marcan 8 elementos (2% del total) para los que no se hallaron clases lo suficientemente específicas ni clases equivalentes:

- *AccionEntrevistaEncuesta*: subclase (tipo de) *AccionEncuesta*;
- *AccionVigilanciaDelTerritorio*: subclase de *AccionPrevencion*;
- *MuestraEncuesta*: subclase de *Grupo*;
- *OficinaDelGabineteDeGobierno*: parte de *Gobierno*;
- *AccidenteAereo* subclase de *Accidente*;
- *AccionReclamaciónDeSoberaníaDeTerritorio*: subclase de *AccionPoliticaExterior*;
- *AdjudicacionContratos*: subclase de *AccionActoAdministrativo*.

4. RESULTADOS

Por lo que respecta a las propiedades, también se identificó una nueva, *alias* (*pen name* en inglés). Como subclase de *nombre*, puede atribuirse a casi cualquier elemento.

Tanto para clases como para propiedades, un análisis en mayor profundidad permitiría definir nuevos elementos a incorporar en ONAPE.

Por último, para cada noticia del tema 1 se calculan exhaustividad, precisión y medida-F exactas (ver Tabla 35). Estos datos muestran tasas elevadas de exhaustividad y precisión, así como de armonía entre las mismas (medida-F). Los datos son aún mayores que en la primera validación, probablemente porque existe mayor homogeneidad de términos entre las noticias analizadas. En todos los casos se obtuvo un resultado superior al **88%** de exhaustividad y precisión, siendo del **100%** para las noticias 1 y 5.

	Noticia 1	Noticia 2	Noticia 3	Noticia 4	Noticia 5	MEDIA
Exhaustividad (exacta)	1	0,88	0,98	0,97	1	0,96
Precisión (exacta)	1	1	1	1	1	1
Medida-F (exacta)	0,5	0,47	0,49	0,49	0,5	0,49

Tabla 35: Exhaustividad, Precisión y medida-F por temas analizados

No procede el cálculo de las medidas cercanas, ya que no se han identificado relaciones de tipo 3 (jerarquía de clase) ni se han tenido en cuenta las de tipo 4 (jerarquía de parte).

4.6. ESPECIFICACIONES DE LAS ONTOLOGÍAS

La creación de la documentación asociada a las ontologías constituye la última etapa de su construcción, de acuerdo con diversas metodologías. Para este proceso se utiliza LODE (*Live OWL Documentation Environment*)⁵¹. Escribir la documentación adecuada requiere un esfuerzo considerable. Por esta razón herramientas que generen la documentación automáticamente y en línea, a partir de una URL o del fichero correspondiente de la ontología son imprescindibles (Peroni, Shotton y Vitali, 2012).

LODE es un servicio web sencillo, orientado a la generación de la documentación de ontologías en HTML para su lectura por seres humanos, tanto si dicha ontología está en formato OWL o RDF. La especificación de ONA está disponible en el anexo B. ESPECIFICACIÓN DE ONA. Por su extensión, la especificación de ONAPE no se ha adjuntado en esta memoria.

*

**

51 LODE: <http://www.essepuntato.it/lode> (último acceso: 05/06/2017)

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Na realidade, a construção de uma ontologia envolve igualmente, como os tesouros, um elevado investimento intelectual, que se pode rentabilizar pelo fato de permitir um tratamento automatizado e tal rigoroso quanto possível do volumen de informação acessível em linha e que contribui para uma análise mais detalhada do dominio sobre o qual se debruça (dos Santos, 2007, p. 67)

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

En este capítulo, se destacan las principales conclusiones alcanzadas, tanto con la investigación desarrollada como con los productos obtenidos, ONA y ONAPE. Esta sección también cuenta con un apartado de trabajos futuros, en el que se indican posibles líneas de investigación, con los modelos como punto de partida. Acaba con un apartado de trabajos publicados durante el periodo de investigación en que se circunscribe esta memoria.

5.1. CONCLUSIONES

Ninguna organización puede pretender que un modelo se ajuste a todos los tipos de cosas, ni para todos los casos, ni siquiera Google (Stuart, 2016, p. 75). No obstante, se ha procurado desarrollar vocabularios generales y reutilizables en diversos contextos, sean noticias u otro tipo de documentos. El marco teórico, basado principalmente en las aportaciones de Fillmore (1967) y García Gutiérrez (2014), se ha revelado útil para el propósito de modelar una ontología general, ONA, a partir de la cual desarrollar una ontología de dominio, ONAPE. Ambos vocabularios han sido validados, primero con las palabras clave empleadas por los propios periódicos como elementos representativos de sus piezas informativas; después, con el análisis del texto completo de una muestra menor de noticias. En ambos casos, se han obtenido altos niveles de la media armónica (medida-F) que evidencian un buen nivel de especialización (exhaustividad) y la adecuación (precisión) de los términos de ONAPE para representar los contenidos analizados.

No obstante, es importante reconocer que una ontología es necesariamente una representación parcial e imperfecta de algo y no un reflejo del complejo mundo real. Escoger [y modelar] una ontología implica la decisión de centrarse en aspectos concretos del mundo real (Stuart, 2016, p. 83). Como todo vocabulario y, específicamente toda ontología, ONAPE no deja de aportar cierta visión del mundo en el dominio analizado, como ya resaltaban Uschold y Grüninger (1996).

Teniendo en cuenta estas cuestiones, a la vista de los resultados de la evaluación de ONAPE, se acepta la hipótesis que se plantea al principio de esta memoria. Esto es, se puede modelar una ontología de dominio político-económico mediante el análisis de una muestra de noticias de medios de comunicación impresa basado en la indización, tanto humana como automática.

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

En cuanto a aspectos concretos de la investigación, se indican las siguientes conclusiones:

1. Los modelados planteados en esta memoria, ONA (ONtología de Acciones) y ONAPE (ONtología de Acciones de Política-Economía) son fruto de la simbiosis de dos áreas de conocimiento ya indicadas anteriormente: Documentación e Informática. Se han aprovechado herramientas desarrolladas por informáticos (e-Termos, Grampal, MetadadosHTML) combinadas con procesos netamente intelectuales: indización humana y análisis de los datos resultantes. Los procedimientos intelectuales han requerido una mayor inversión de tiempo, vinculada con la complejidad del dominio político-económico, más aún en un contexto internacional de medios de comunicación. Esto se debe a que en cada país existe una realidad distinta, que después es descrita por un medio de documentación y, finalmente, por personas, por lo que es difícil interpretar qué se dice y qué se quiere decir, incluso para otro ser humano. La experiencia previa y los conocimientos adquiridos antes y durante la investigación han sido esenciales para este análisis.
2. En relación con la fundamentación metodológica, esta investigación contribuye a visualizar la relación entre distintos vocabularios (taxonomías, tesauros, ontologías), desarrollados con vistas a la representación del contenido de los documentos (representación semántica). La ontología, siguiendo a Codina y Pedraza-Jiménez (2017), entre otros, es presentada como una evolución de las herramientas que han ido apareciendo a lo largo de la historia, de acuerdo con las necesidades de recuperación de información de cada época. Este vínculo es importante, ya que permite acercar dos posturas que, a la vista de la bibliografía utilizada, se siguen separando: vocabularios desarrollados en el área de documentación, vocabularios del ámbito informático.
3. Por lo que respecta a las metodologías de modelado de ontologías y al objetivo 2 de esta tesis, como indican Mendonça y Almeida (2014) y Silva (2008) existen distintas aportaciones. Sin embargo, ninguna metodología se ha alzado como estándar, por lo que se ha optado por seguir las pautas recogidas en las referencias indicadas. Esta perspectiva se adapta, en términos generales, a la creación de cualquier otro vocabulario y, aunque no se ha seguido la nomenclatura de los autores, queda más o menos patente a lo largo de la memoria. Al definir la metodología se han empleado un lenguaje y grado de descripción que facilitan su entendimiento y posible reutilización para el modelado de ontologías de dominio, así como para la descripción y recuperación de documentos con ONAPE. Los vocabularios no siempre van a ser desarrollados por documentalistas y/o informáticos/as.

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

4. Con la indización humana y la semiautomática, empleadas para la asignación, en el primero caso, y la extracción, en el segundo, de términos de indización con los que dotar de clases y relaciones a ONAPE, se ha conseguido un buen número de elementos. Esta investigación es consistente con algunos aspectos ya señalados por García Gutiérrez (2014) sobre la capacidad representativa de los titulares: muchas veces «están adornados de una implacable retórica dentro de una tipología textual muy extensa», entre las que se puede destacar el uso de oraciones impersonales, oraciones transitivas interrumpidas u oraciones con omisión verbal. Por esta razón, la indización humana facilita su interpretación. Incrementar el tamaño de los textos de análisis, por otro lado, puede proporcionar una mayor cantidad de elementos potenciales. El ruido que esta nueva aplicación aporte en el proceso debe ser analizado para evaluar sus costes en relación con la calidad del producto. El Procesamiento del Lenguaje Natural (PLN), que ha sido utilizado tanto para extraer términos para una ontología o tesaurus, también se puede emplear para la indización de documentos. La colaboración con profesionales de la Informática es necesario para ello (ver 5.2. TRABAJOS FUTUROS).
5. En lo que tiene que ver con el objetivo 2 de la tesis, se han identificado precedentes de indización automática en documentos periodísticos, por ejemplo García et al. (2006) o Maidel et al. (2010). Sin embargo, la mayoría utilizan taxonomías generales de la IPTC, como Troncy (2008), y no descienden a dominios concretos ni ponen el foco en las acciones descritas en las noticias. Por lo que respecta a política-economía, se han encontrado trabajos relativos a ontologías especializadas, como Ortiz (2011) o Daquino et al. (2014), pero se centran en aspectos distintos o abordan procedimientos diferentes a los planteados. En el anexo A.1. HERRAMIENTAS REUTILIZADAS POR LOS VOCABULARIOS se describen con mayor detalle las características de algunas de las ontologías analizadas, generales y especializadas en prensa y política-economía, de cara a una reutilización posterior.
6. En relación con los esquemas de metadatos y al objetivo 3, se han analizado tanto esquemas generales como aplicados a los documentos propios de los medios de comunicación impresa y al dominio política-economía. Esto ha permitido el acercamiento a los metadatos de contenido (semántico). En Baños-Moreno et al. (2015) se destaca, en el ámbito de la prensa escrita, la convivencia de diversos estándares de metadatos. Tanto esquemas de metadatos como ontologías tratan de hacer frente a los mismos problemas: la fugacidad de la información en las redes, la multiplicación de la información y de los vocabularios diseñados para su mejor control. En este sentido, en Baños-Moreno et al. (2017) se pone en evidencia que no sólo hay un problema de cantidad, sino de calidad de la descripción de

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

información, con prácticas de representación no sistemáticas ni compartidas, lo que dificulta la reutilización de información. Por esta razón, se ha previsto completar el mapeo de ONA y ONAPE con los vocabularios analizados (ver 5.2. TRABAJOS FUTUROS). Así, en un futuro se reemplazarán elementos de las ontologías por aquellos utilizados como fuente para la definición de clases y propiedades. Además, se pondrían a disposición en un alojamiento permanente de la Universidad de Murcia, con su documentación asociada. De esta forma se pretende atender los principios de la web semántica de facilitar el acceso a los vocabularios para su utilización en distintos contextos.

7. También relacionado con el objetivo 3 de esta tesis, *MetadadosHTML*, software *ad hoc* para la extracción de metadatos del código fuente de noticias ha resultado útil para la obtención de metadatos de tipo descriptivo con vistas al análisis de los mismos. Sin embargo, destaca la dependencia de los cambios en los códigos fuente. Si un periódico redefine alguna de sus etiquetas meta o comienza a utilizar un nuevo esquema, el software no puede aprovecharlas, lo que obliga a una continua revisión. Además, destaca la dificultad de recoger todos los atributos mostrados en los códigos fuente, ya que no siempre están disponibles. De acuerdo con esto, se percibe la necesidad de definir, por parte de los periódicos, una política de indización alineada con estándares de metadatos. Es evidente que la simple adopción de metadatos no es suficiente para la difusión y accesibilidad de los documentos.
8. En cuanto a la actualización del vocabulario de representación, más aún en un dominio tan amplio y complejo como es política-economía, ONAPE debe ser revisada. Implica la inclusión de nuevos aspectos no contemplados hasta el momento y, quizá, la supresión de otros. Para ello, es necesario analizar nuevos *corpora* de noticias (u otros documentos) y, como destaca Hlomani (2014), adecuarla a cada periodo de análisis. Sólo de esta forma se puede asegurar la validez de ONAPE en el tiempo. Es decir, que responda adecuadamente a las necesidades de descripción de nuevas realidades. Tal como se ha planteado la construcción de ambas ontologías, así como su evaluación, contemplada como objetivo 4 de la tesis, esta investigación constituye el punto de partida de un mayor desarrollo que tenga en cuenta no sólo su actualización, sino una evaluación en mayor profundidad. Para ello, se define como propuesta de investigación la actualización de la ontología mediante aprendizaje basado en la experiencia (ver 5.2. TRABAJOS FUTUROS).

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

9. Finalmente, la evolución de la web sintáctica (Romero Llop, 2007) a la web semántica es un camino dinámico en el que aparecen y continuarán surgiendo nuevas técnicas y vocabularios orientados al tratamiento de la información. Estas herramientas facilitarán el mercado semántico del cada vez mayor volumen de contenidos disponibles en Internet. Para ello, los/as informáticos/as desarrollan técnicas y algoritmos. Sin embargo, es necesaria la participación activa de los/as profesionales de la información, como documentalistas, archiveros/as, bibliotecarios/as, etc., que cuentan con experiencia y formación para mejorar los resultados aportados por la informática. Tennis y Calzada-Prado (2007) ya destacaban el campo de posibilidades que se abría para los/as profesionales de Biblioteconomía y Documentación a este respecto. Documentalistas y profesionales afines son quienes se enfrentan diariamente a los problemas de acceso a la información, ya sean por una descripción de la información insuficiente o por carencias en la alfabetización digital. Desde un punto de vista formativo, es necesario incrementar la carga lectiva de los grados de Información y Documentación en tecnologías de la web semántica. Entender las estructuras de los lenguajes de marcado, como HTML, XML, RDF u OWL, puede contribuir a un mejor análisis de los vocabularios en lenguaje formal y sus posibilidades. Aprender a utilizar las diversas herramientas de procesamiento automático de la información, además, puede reducir las tareas mecánicas, dedicando ese tiempo a actividades intelectuales con mayor aporte a la investigación.

5.2. TRABAJOS FUTUROS

En esta sección, se recogen propuestas de trabajos futuros, relacionados con ONA y ONAPE.

MAPEADO CON VOCABULARIOS REUTILIZADOS

Como se ha indicado anteriormente, esta propuesta consiste en completar el mapeo, tanto de ONA como de ONAPE, con elementos de vocabularios descritos en 4.4.3. MAPEO DE VOCABULARIOS. Este proceso es iniciado en esta investigación mediante la inclusión de propiedades tipo anotación (*isDefinedBy*), especificando la fuente de definición de determinadas clases y propiedades. Este futuro trabajo pondría el foco en la sustitución de estos componentes por elementos importados del vocabulario de origen.

ESPECIALIZACIÓN EN SUBDOMINIOS

El dominio política-economía es amplio, por lo que cabe plantearse el desarrollo de subdominios de interés, en función de las necesidades informativas que puedan surgir en el futuro. Por ejemplo, las actividades codificadas en CNAE (Codificación Nacional de

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Actividades Económicas)⁵² pueden ser útiles para aportar más detalles por lo que respecta a las acciones *AccionActividadEconomica*, creando nuevos elementos y/o para completar definiciones, descripciones y relaciones de sus subclases.

Otro aspecto más interesante tiene que ver con la corrupción, política y/o económica. No son pocas las noticias que diariamente informan de nuevos casos de corrupción, tanto en España como en otros países. Un mejor control de la información de este tema puede facilitar la toma de decisiones de los/as ciudadanos/as, por ejemplo, en los procesos electorales. También puede ser útil para otro tipo de usuarios. Así, periodistas y documentalistas de un medio de comunicación podrían recuperar información más precisa sobre un caso concreto mediante el marcado de documentos con elementos de un subdominio de ONAPE más especializado. Tal es el interés que, de hecho, ya existen precedentes de análisis de documentos relacionados con esta temática mediante ontologías, como Corrupteca, descrita brevemente en 4.4.1. PRESELECCIÓN DE VOCABULARIOS.

ACTUALIZACIÓN DE ONAPE MEDIANTE EL ANÁLISIS DE NUEVOS CORPORA

Consiste en la inclusión de nuevos elementos que permitan modular las acciones, estableciendo grados de ejecución de éstas o bien especificando con mayor detalle qué ocurre. Por ejemplo, se puede plantear una noticia, relativamente frecuente en los periódicos: suben las ventas de un producto x. La acción que se ejecuta, en sí, no es más que la venta del dicho producto o, más exactamente, la venta repetida de dicho producto. Para contextualizar mejor esa noticia, podría emplearse una propiedad, *tieneModulador*, que vinculara una acción con una subclase de *ModuladorDeAccion* (ver Figura 75). Así, en el ejemplo anterior, una *AccionVenta* podría tener *Subida* como modulador de dicha acción.

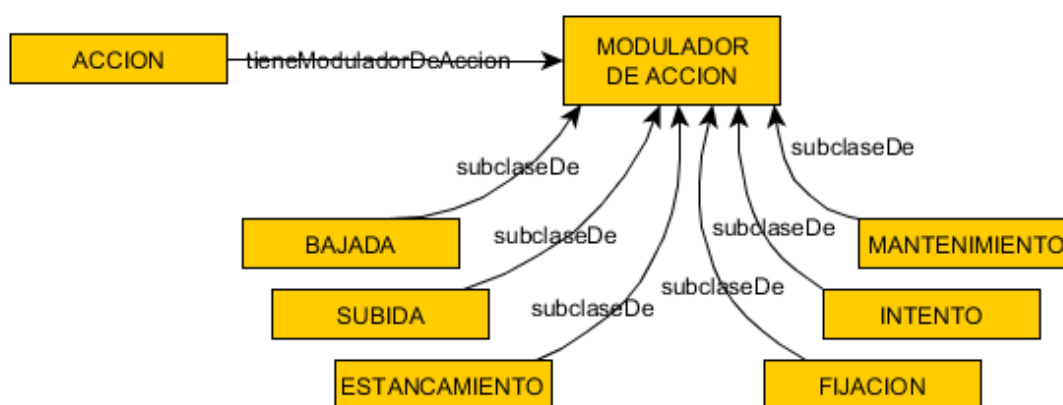


Figura 75: Relación entre Accion y ModuladorDeAccion

52 CNAE: <http://www.cnae.eu/> (último acceso: 18/05/2017).

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Estas subclases pueden definirse de la siguiente forma:

- **Bajada:** indica la caída o disminución de algo. Por ejemplo, la bajada o caída del número de usuarios de un producto. Términos que reflejan la idea de caída son Reducción, Disminución, Declive, Decrecimiento, Caída, Desplome y Descenso.
- **Estancamiento:** permite expresar el mantenimiento de las cantidades o cifras de ejecución de alguna acción.
- **Fijación:** clase para expresar el establecimiento o fijación de algún elemento en algún tipo de Acción, como el Precio de un Producto el salario de trabajadores/as.
- **Intento:** intento y fracaso de conseguir algo a través de la ejecución de una Acción.
- **Mantenimiento:** clase por la que se expresa el mantenimiento de alguna condición o situación.
- **Subida:** clase por la que se expresa el aumento de algo, en algún tipo de acciones. Se puede utilizar, por ejemplo, para expresar el crecimiento de las ventas de un producto. Términos que reflejan la misma idea son Aumento, Crecimiento, Expansión e Incremento. También Escala y Escalada se emplean reflejando el incremento o subida de algo, por ejemplo, Escalada de Tensión.

Ahora bien, el mantenimiento y sostenibilidad de las ontologías son complicados. Su creación ya conlleva un consumo de tiempo elevado y persiste si se quiere mantener disponible y actualizada (Stuart, 2016, p. 125). Por ello, se plantea como trabajo futuro un proyecto de instanciación (y actualización) de ONAPE mediante el aprendizaje basado en la experiencia. Esta técnica prevé la automatización de la instanciación de la herramienta mediante una etapa previa de marcado semántico humano. De esta forma, el sistema aprende no sólo qué términos son instancias, sino con qué estructuras suelen aparecer ligadas. Por ejemplo: determinante + sustantivo + adjetivo + verbo.

Por lo que respecta a los *corpora* de análisis, es posible recurrir a varias opciones en el marco de las noticias de prensa escrita: 1) ampliar el tamaño del texto a analizar ya que, según Guerrillot (2006), el asunto principal de la noticia se encuentra entre las 100 primeras palabras, uno o dos párrafos; 2) construir macroproposiciones, en el sentido en que las explica García Gutiérrez (2014): «síntesis textual elaborada a partir de la formulación de las 12W a cada una de las noticias de la muestra que consiste en “reescribir” (y, en su caso, sustituir) y enriquecer el título en un mismo enunciado canónico del tipo: sujeto-acción-predicado con los datos que el propio texto original explícitamente proporcione». El primero facilitaría la aplicación de técnicas de PLN, mientras que el segundo requeriría de una etapa intelectual, en la que se identificaran los elementos y estructuras necesarias para construir esas macroproposiciones.

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Otra posibilidad es el recurso a los contenidos producidos en el contexto de los medios de comunicación, pero no por los periodistas, sino por los lectores, tales como *posts*, comentarios en noticias, etc. Estos, convertidos en *prosumidores* (Toffler, 1980) participan activamente en la generación de contenidos, gracias a herramientas como blogs, comunidades o comentarios, dentro y fuera de las webs de los periódicos. De hecho, *The New York Times*, *The Guardian* y la BBC, entre otros, han comenzado a publicar algunos de sus contenidos directamente en la red social (Jiménez Cano y Abad Liñán, 2015). Su control y recuperación resultarán fundamentales para responder adecuadamente a las necesidades informativas que, como ciudadanos/as, tenemos. Para ello, conviene conocer los niveles de interactividad de lectores/as con respecto a los distintos elementos de las páginas web de los medios de comunicación impresa. En este sentido, un análisis inicial es realizado en (Baños-Moreno, Pastor-Sánchez y Martínez-Béjar, 2017).

Por último, también se plantea recurrir a las entrevistas a expertos, tanto de los dominios de política-economía, como de periodismo y de documentación. El propósito es mejorar la herramienta: ¿qué problemas encuentran los investigadores y profesionales para recuperar documentos?, ¿qué les gustaría encontrar en un hipotético sistema de recuperación de información?

Una cuestión adicional tiene que ver con la inclusión de sinónimos de clases y relaciones, enriqueciendo así las posibilidades de marcado semántico de documentos. Ello facilitaría la identificación de términos que, siendo diferentes, hacen referencia a un mismo concepto. Para ello, se analizarían qué fuentes son ricas en terminología, como Eurovoc, y ontologías lingüísticas, como Wordnet. Posteriormente, se definirían las relaciones entre los términos de ONAPE y los de estos vocabularios.

MEJORA DE LA EVALUACIÓN DE LAS ONTOLOGÍAS

Otro aspecto en el que trabajar es la profundización del análisis de las ontologías con las medidas de evaluación utilizadas: exhaustividad, precisión y medida-F. Para ello, se plantea emplear una muestra mayor de noticias que facilite el análisis de datos acumulados. Asimismo, es posible el uso de herramientas de análisis especializadas.

MEJORA DE METADADOSHTML

Los resultados iniciales evidencian la necesidad de establecer un orden de preferencia en la extracción de valores de datos a partir de unos esquemas u otros. Para ello, MetadadosHTML será modificado teniendo en cuenta la frecuencia de uso de los esquemas y el grado de normalización de las prácticas de cada medio. La Tabla 1 indica el orden de extracción de palabras clave empleado.

Los códigos de la columna 3 se indican en la Tabla 2 del apartado 4. RESULTADOS.

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Orden	Línea de código	Medios	Nota
1	<meta name="keywords" content="">	101; 121; 123; 202; 311; 401; 601; 602; 603; 701; 711; 721	Para 201 se utiliza 4, 5, 6 y 7 Campo repetible en 701
2	<meta name="news_keywords" content="">	122; 731; 801	Para 201 se utiliza 4, 5, 6 y 7
3	<meta property="article:tag" content="" />	301; 901	Campo repetible
4	<meta name="des" content="" />	201	Palabras clave
5	<meta name="org" content="" />	201	Organizaciones
6	<meta name="per" content="" />	201	Personas
7	<meta name="geo" content="" />	201	Lugares

Tabla 1: Información sobre el metadato "palabras clave" para MetadadosHTML

El propósito es asegurar la correcta extracción de valores de datos con el menor número de procesos. La información extraída será utilizada para la construcción de una ontología de dominio, facilitando la definición de clases, subclases y relaciones.

5.3. PUBLICACIONES CIENTÍFICAS

Se encabeza, en color azul aquellos trabajos que ya han sido publicados. Aquellos en color morado se encuentran en proceso de revisión y/o publicación.

ARTÍCULOS DE REVISTA

Autores/as	BAÑOS-MORENO, M.J; Pastor-Sánchez, J.A; Martínez-Béjar, R.
Título	<i>Interactivity features of online newspapers: from a facsimile model to a multimedia one</i>
Publicación	Anales de Documentación
Volumen, nº	20, 2
Fecha de pub.	2017
Resumen	The purpose of this article is to know the current interactivity options in online newspapers. To do this, we analyse the concept of (structural) interactivity and establish a measuring tool based on some previous methodologies. This model includes a classification of newspapers into stages of development. In this way, we provide a tool, which allows media firms to know the degree of interactivity in newspapers, assessing what dimensions and parameters are being incorporated. It was applied to a sample of 21 online newspapers. Results show poor implementation of participation and customization options. There is a niche to be covered by sections of user-generated content and closer collaboration journalists-citizens. It will require a real adaptation to the new possibilities of interaction with readers at different levels of involvement and participation. The majority of newspapers are classified in a digital stage and some of them are close to the multimedia phase.
Enlace	http://revistas.um.es/analesdoc/article/view/282401
IF	IF SJR (2016): 0,101 (Q4)

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Autores/as	BAÑOS-MORENO, M.J; Felipe, E.R.; Pastor-Sánchez, J.A; Lima, G.; Martínez-Béjar, R.
Título	<i>Análisis de metadatos de noticias para la extracción de información de código fuente. El software METADADOSHTML</i>
Publicación	Information Research – An International Electronic Journal
Volumen, nº	22, 1
Fecha de pub.	2017
Resumen	<p><i>Introducción.</i> Los objetivos de este trabajo son determinar qué esquemas se utilizan para título, resumen, palabras clave, autoría y periódico en prensa; conocer qué pautas siguen los periódicos en la implementación de dichos esquemas; y averiguar cómo esto afecta a la extracción de información.</p> <p><i>Metodología.</i> Para ello, se define una muestra de diarios y se analiza su código fuente, identificando esquemas utilizados y patrones de uso. Esto permite extraer valores de dato utilizando la aplicación MetadadosHTML.</p> <p><i>Resultados.</i> Se han detectado esquemas estándar, ad hoc y propios de los periódicos. Se han hallado diversas prácticas, como valores agrupados en una misma línea de código o por separado; ruido en un valor y errores al referir los nombres de los atributos de esquemas estándar. Estos problemas afectan a la extracción de datos basada en esquemas de metadatos y metadatos en MetadadosHTML.</p> <p><i>Conclusiones.</i> Es necesario avanzar en el uso de esquemas estándar, como Dublin Core o schema.org, favoreciendo la implantación de estos (u otros) en los códigos fuente de noticias. También resulta imprescindible la adopción de buenas prácticas al explicitar datos y valores de datos. Sólo así es posible evolucionar en la interoperabilidad entre sistemas y en la recuperación y reutilización de información.</p>
Enlace	http://www.informationr.net/ir/22-1/paper740.html
IF	JCR (2016): 0,574 (Q3); IF SJR (2016): 0,3 (Q2)
Autores/as	RUIZ-MARTÍNEZ, J.M; BAÑOS-MORENO, M.J.; MARTÍNEZ-BÉJAR, R.
Título	<i>La nomenclatura UNESCO: Evolución, alcance y reutilización en clave ontológica para la descripción de perfiles científicos</i>
Publicación	El Profesional de la Información
Volumen, nº	23, 4
Páginas	383-392
Fecha de pub.	2014
Resumen	<p>En 1974 la Unesco publicó la Nomenclatura Normalizada Internacional para los campos de Ciencia y Tecnología. Desde entonces se ha utilizado para la clasificación de la actividad científica en España. Con el objetivo de conocer mejor sus características, evolución y alcance se ha realizado un estudio bibliográfico y un análisis de su utilización en grupos de investigación universitarios y en otros ámbitos nacionales e internacionales. A partir de los datos obtenidos se ha diseñado una ontología basada en la combinación de la Nomenclatura y palabras clave.</p> <p>El resultado es un modelo ontológico adaptado a las tecnologías de la web semántica, que proporciona mayor expresividad que la Nomenclatura y soluciona problemas como la obsolescencia y la ausencia de ciertos temas, facilitando la clasificación de la actividad científica e incrementando su utilidad.</p>
Enlaces	http://recyt.fecyt.es/index.php/EPI/article/download/epi.2014.jul.06/16967
IF	JCR (2016): 1,063 (Q3); IF SJR (2016): 0,541 (Q1) JCR (2014): 0,356 (Q4); IF SJR (2014): 0,451 (Q2)

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

Autora	BAÑOS-MORENO, M.J.
Título	<i>Fuentes para la actualización de macro-tesauros: Noticias de divulgación científica</i>
Publicación	Cuadernos de Gestión de Información
Volumen, nº	3, 1
Páginas	13-24
Fecha de pub.	2013
Resumen	<p>Los tesauros son herramientas de organización del conocimiento necesarias para el control de la información. Muchos no se actualizan con la necesaria frecuencia, reduciendo considerablemente su utilidad. Este trabajo tiene como objetivos, analizar el grado de actualización de los tesauros de la Unesco y Unión Europea (Eurovoc); y determinar la capacidad de las noticias de divulgación científica de ser usadas para efectuar esa renovación.</p> <p>Para ello se han extraído palabras clave a partir de titulares de artículos de divulgación científica, publicados en prensa digital. A continuación, mediante técnicas de Recuperación de Información (N-gramas y Apache Solr) se buscaron equivalencias con los tesauros mencionados anteriormente.</p> <p>Los datos obtenidos permiten confirmar que ninguno de los macro-tesauros debe utilizarse para describir noticias de divulgación científica, ya que buena parte de las palabras clave extraídas no se recogían en estos. Sin embargo sí pueden emplearse como base para la construcción de otros vocabularios de organización del conocimiento. Por otro lado, una revisión más frecuente es necesaria y obligatoria. La inclusión de nuevos términos y la redefinición de las relaciones entre los ya existentes incrementarían incuestionablemente la calidad de las herramientas analizadas. Las noticias de divulgación resultan adecuadas para dicha actualización, constituyendo auténticos yacimientos de conceptos y términos.</p>
Enlace	http://revistas.um.es/gesinfo/article/view/207731 / http://eprints.rclis.org/21174/

Tabla 2: Artículos publicados

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

CONTRIBUCIONES A CONGRESOS

Autora	BAÑOS-MORENO, M.J.
Título	<i>Fuentes de búsqueda de ofertas de empleo en el sector privado para profesionales de la información en España. Una experiencia personal</i>
Lugar	Madrid
Editor	Universidad Complutense de Madrid y Universidad de Brasilia
Congreso	V Seminario Hispano Brasileño de investigación en Información, Documentación y Sociedad
Fecha realiz.	14-17 Noviembre 2016
Resumen	<p>El objetivo de este trabajo es ofrecer un marco general de fuentes de información sobre ofertas de empleo para el profesional de la gestión de información, esto es, una guía básica para la búsqueda de empleo activa y eficiente, mostrando como producto final la descripción y clasificación de estos recursos y una serie de recomendaciones para su uso. Para ello, se han analizado algunas publicaciones especializadas para conocer qué denominaciones existen para el/la profesional de la información. Después se analizan y describen una serie de recursos en línea para la búsqueda de ofertas de trabajo en el sector de la información y documentación, indicándose por qué fuentes comenzar para tener acceso a una buena parte de estas ofertas.</p> <p>También se incluyen recursos habitualmente conocidos por los/as profesionales, a los que se acaba llegando cuando la persona entra en el mundo formativo y/o laboral. Posteriormente, estos recursos son clasificados, utilizando diversos criterios que facilitan su mejor conocimiento. Mediante la observación de los recursos y de las ofertas publicadas, y de la postulación a estas, se muestran algunas impresiones sobre el funcionamiento y características de los recursos analizados. Finalmente, a partir de la experimentación con los recursos, utilizando algunas de las denominaciones para el/la profesional de la información indicadas, se recoge una serie de recomendaciones de uso de los recursos, que dan preferencia a ciertos recursos y a las notificaciones y alertas por correo. De esta forma, se evita la duplicidad de tareas y la pérdida de tiempo en tareas que pueden ser automatizadas, ganando en eficiencia.</p>
Enlace	http://eprints.rclis.org/31636/
Autores/as	BAÑOS-MORENO, M.J.; FELIPE, E.R.; PASTOR-SÁNCHEZ, J.A.; MARTÍNEZ-BÉJAR, R.; LIMA, G.
Título	<i>Metadatos en noticias: un análisis internacional para la representación de contenidos en periódicos</i>
Editores	Rodríguez Muñoz, J.V.; Gil-Leiva, I.; Díaz Ortuño, P.; Martínez Méndez, F.J.
Lugar	Murcia
Editor	Facultad de Comunicación y Documentación, Universidad de Murcia
Congreso	II Congreso ISKO España y Portugal / XII Congreso ISKO España
Fecha realiz.	19 y 20 de noviembre de 2015
Páginas	290-303
Resumen	<p>Los periódicos trabajan con un gran volumen de información que necesita ser descrita adecuadamente. Para ello, las etiquetas “title”, “keywords” y “description” son muy utilizadas en el código fuente de las noticias online. Sin embargo, estas no resultan suficientemente descriptivas. Así, surgen estándares de metadatos, con el fin de facilitar la interoperabilidad y profundizar en la descripción. Actualmente, las etiquetas HTML y diversos estándares conviven en el sector periodístico, con diversos grados de implantación.</p>

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

	<p>Se analiza el código fuente de diarios internacionales de información general y se lleva a cabo una profunda revisión bibliográfica sobre estándares de metadatos. El propósito es conocer qué estándares existen, y evaluar su uso en los códigos fuente de una muestra de periódicos. Para ello se identifican los metadatos de contenido semántico de los códigos fuente. Además se desarrolla el software MetadadosHTML.</p> <p>Como conclusiones destacan la gran distancia entre los estándares recogidos en la bibliografía y los mostrados en los código fuente. En el primer caso, los más referenciados son los formatos NewsML y NITF, implementados por algunos medios y agencias de prensa, al menos a nivel interno. Por el contrario, en el código fuente los más habituales son schema.org y dos esquemas para mostrar información en redes sociales, Open Graph Protocol (usado por Facebook) y Twitter Cards. Esto, evidencia la convivencia de diversos estándares de metadatos en el ámbito de los medios de comunicación y pone de relieve la falta de uniformidad en su uso. Para alcanzar el ideal de interoperabilidad de los contenidos, es preciso utilizar tecnologías de la Web Semántica. En este sentido, se debería tender a definir ontologías o vocabularios RDF para las diferentes propuestas analizadas.</p>
Enlace	http://eprints.rclis.org/28589/
ISBN	978-84-608-3558-5
Autores/as	BAÑOS-MORENO, M.J.; PASTOR-SÁNCHEZ, J.A.; MARTÍNEZ-BÉJAR, R.
Título	<i>Propuesta de actualización de macro-tesauros a partir de noticias de divulgación científico-tecnológica</i>
Editores	RIBEIRO, Fernanda y CERVEIRA, Maria Elisa (org.)
Lugar	Porto (Portugal)
Editor	Faculdade de Letras da Universidade do Porto - CETAC.MEDIA
Congreso	I Congreso ISKO España y Portugal / XI Congreso ISKO España
Fecha realiz.	7-9 November 2013
Páginas	99-112
Resumen	<p>Los tesauros son herramientas de organización del conocimiento necesarias para el control de la información, más aún en el mundo de la información digital. Sin embargo, muchos de ellos adolecen de una falta de actualización que reduce considerablemente su utilidad. Este trabajo tiene precisamente como objetivo analizar el grado de actualización de dos de los tesauros más importante, el de la UNESCO y el de la Unión Europea (Eurovoc). Así, para la descripción de artículos de prensa, se procedió a la extracción de términos descriptivos de contenido a partir de titulares de artículos de divulgación científico-tecnológica publicados en prensa digital. A continuación mediante técnicas de Recuperación de Información se buscaron equivalencias con los tesauros mencionados anteriormente. Los resultados obtenidos muestran un nivel de equivalencia exacta o cercana que ronda el 50%. Este porcentaje se aproxima al 75% considerando equivalencias jerárquicas y asociativas. Este dato permite confirmar que ambos macro-tesauros pueden ser la base para elaborar otros vocabularios. En el caso concreto de su aplicación para la indización de noticias de divulgación científico-tecnológica, Eurovoc es ligeramente mejor que el Tesauro de la UNESCO, puesto que los términos y relaciones están más actualizados. El nivel de equivalencia exacta o cercana indica que las noticias de divulgación científico-tecnológica constituyen una fuente adecuada para la actualización de tesauros bien para la inclusión de nuevos términos o para la redefinición de las relaciones entre estos</p>
Enlace	http://eprints.rclis.org/20684/
ISBN	978-989-8648-10-5

Tabla 3: Contribuciones a congresos

5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES

CAPÍTULOS DE LIBRO

Autora	BAÑOS-MORENO, M.J.
Título	<i>Evolución de los instrumentos para la representación del conocimiento: de las listas de términos a las ontologías</i>
Publicación	Bibliotecas digitais: novas tendências na navegação em contexto
Lugar de pub.	Belo Horizonte, Brasil
Editorial	Universidade Federal de Minas Gerais (UFMG)
Fecha de pub.	2017 (en prensa)
Resumen	Se presentan y describen cuatro de los vocabularios de representación del conocimiento más habituales: listas de términos, taxonomías, tesauros y ontologías.
Páginas	---
ISBN	---
Autores/as	HERNÁNDEZ, P.; GARCÍA DÍAZ, I.; BAÑOS-MORENO, M.J., PALAZÓN-CEBRIÁN, A.; MARGALEF, J.M.I
Título	<i>Quiebra y crisis de modelo de la política pública de cultura en la Región de Murcia</i>
Publicación	FORO CIUDADANO DE LA REGIÓN DE MURCIA. El otro estado de la Región. Informe 2014
Lugar de pub.	Murcia
Editorial	Diego Marín / Foro Ciudadano de la Región de Murcia
Fecha de pub.	2014
Resumen	Se presenta la situación actual (2014) de la cultura en la Región de Murcia
Páginas	229-250
ISBN	978-84-16296-28-6

Tabla 4: Capítulos de libro publicados durante el periodo de investigación

PRESENTACIONES DE LA INVESTIGACIÓN

Autora	BAÑOS-MORENO, M.J.
Título	<i>Modelagem de uma ontologia de domínio baseada na Gramática de Casos de Fillmore</i>
Lugares de exposición	Grupo de Pesquisa MHTX, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil Grupo de Pesquisa Tecnologias em Ambientes Informacionais, Universidade Federal de São Carlos, São Paulo, Brasil
Fecha de exposición	01/04/2016 y 28/03/2016

Tabla 5: Presentaciones de esta investigación

*

**

RESUMO EM PORTUGUÊS

Neste capítulo se resume a pesquisa desenvolvida em português com a exceção da seção das conclusões e trabalhos futuros, que são traduzidos para esta língua. Desta forma, se cumpre com o artigo 31b) da [Resolución del Rector de la Universidad de Murcia \(R-310/2015\), de 15 de mayo, por la que se ordena la publicación en el Boletín Oficial de la Región de Murcia de la modificación del Reglamento de Doctorado](#), onde se diz o que segue: *parte da tese doutoral, pelo menos o resumo e as conclusões, devem ser redigidas e apresentadas numa língua habitual para a comunicação científica no seu campo de conhecimento, diferente das línguas oficiais da Espanha.*

1. INTRODUÇÃO

A realidade político-econômica é, provavelmente, uma das mais interessantes. As decisões de quem nos governam, os efeitos das ações das pessoas para as pessoas, são fatos que devem ser analisados para saber o que acontece no mundo. A importância deste domínio na política e economia pode ser traduzida numa necessidade: informação organizada e acessível, que melhore a contextualização e a compreensão do que temos ao redor.

Parte desta necessidade é resolvida pelos jornais. O consumo de informação é uma das maneiras de agir para conectar com a realidade (Hernández-Serrano et al., 2017). Pela sua influência, tanto na vida privada quanto na pública, as informações sobre Política e Economia encontram-se entre as mais destacadas e prolíferas nos meios de comunicação (Moreira et al., 2013; Río Cabrerizo, 2014).

Os avanços tecnológicos e, sobretudo a Internet, multiplicaram a quantidade de informação disponibilizada, bem como as fontes para o seu acesso. Mas isso, não é uma questão atual, pois nos anos 50 e 60 o crescimento descontrolado da informação já era um problema analisado. Hoje, esse crescimento da informação e o seu compartilhamento virou exponencial (Stuart, 2016, pp. 2-4), gerando mais desorganização. Por outro lado, a Web foi projetada por humanos e para humanos, tornando-a uma imensa biblioteca caótica, em que a extração do conhecimento é difícil e de preço elevado (Castells et al., 2006; Díaz Nosty, 2013; Mannens et al., 2013; Nies et al., 2012). Alguns autores chamam este problema de *infoxicação* (Cornella, 2010) ou sobressaturação de informação (Toffler, 1973, pp. 247-250), ou seja, a incapacidade das pessoas para gerenciarem o atual volume de informação disponibilizada.

As bibliotecas, arquivos e outros tipos de centros especializados têm desenvolvido métodos e ferramentas para resolver esta situação, mas não tem sido suficiente, já que o problema dos documentos web está relacionado com a ausência de semântica compreensível pela máquina. É preciso desenvolver ferramentas que assistam na recuperação de informação e que aumentem a precisão sem ter um impacto elevado na exaustividade (Stuart, 2016, p. 2).

Há algum tempo surgiu o que Berners-Lee, Hendler e Lassila denominaram web semântica, uma extensão da web atual na qual a informação disponível tem um significado bem definido e cujo objetivo é que dados definidos e interligados sejam utilizados com mais efetividade na descoberta, automatização, integração e reutilização de informação por múltiplos aplicativos (Hendler et al., 2002).

Nos anos noventa e como parte da web semântica, as ontologias emergiram como tema de muitas pesquisas. Trata-se de um tipo de ferramenta para a gestão do conhecimento que, devido às suas possibilidades de interoperabilidade e de representação de domínios, permite a compreensão compartilhada destes (tanto por pessoas quanto por máquinas), tornando-se popular em diversas comunidades científicas. Na área de Ciência da Informação, segundo Codina y Pedraza-Jiménez (2011), é possível que acabe por se impor como vocabulário para descrever e recuperar informações. Deste modo, as ontologias contribuem para solucionar os problemas levantados: a multiplicação da informação, o seu compartilhamento e a ausência de semântica compreensível pela máquina.

Embora já existam ontologias no domínio da política-economia e, sobretudo, de meios de comunicação, muitas delas se restringem a protótipos, não ficaram disponibilizadas para o seu acesso, só tem relações hierárquicas ou focam-se em subdomínios muito concretos.

Esta tese tem como motivação a preocupação pela situação político-econômica internacional, período de muitas guerras, terrorismo, perdas de direitos sociais, prêmios de risco, quedas de vários governos, etc. Diariamente, as pessoas recebemos múltiplos estímulos informativos, mas as possibilidades de busca e recuperação de informação são insuficientes. As ontologias são úteis para, como sujeitos políticos, ter uma visão completa dos acontecimentos, bem como fornecer a população uma opinião diferenciada. Por outro lado, os vocabulários controlados são muito presentes na formação dos profissionais da área da Ciência da Informação e como tal, temos de nos aproximar à Web Semântica (Stuart, 2016, p. 47).

1.1. OBJETIVOS

Esta tese propõe a modelagem de uma ontologia para a representação de ações no domínio político-econômico com vistas a facilitar a compreensão dos acontecimentos do mundo real. Desta forma, pretende-se contribuir para a área da Ciência da Informação com o desenho de uma ferramenta que forneça e melhore a descrição de documentos do referido domínio. Também foi considerado que tal modelagem pode ser útil para facilitar a organização e a recuperação da informação.

Este objetivo é dividido nos seguintes alvos específicos:

1. Conhecer metodologias de modelagem de ontologias já existentes para a partir delas, definir as etapas para a modelagem e construção da ontologia proposta. Com esses passos, se propõe a definição de um marco teórico de análise baseado na gramática de casos para a representação de ações com ontologias;
2. Conhecer ontologias gerais, do domínio analisado e de meios de comunicação, bem como avaliar as possibilidades de reutilização delas no modelo proposto;
3. Identificar quais esquemas de metadados são utilizados na imprensa para representar o conteúdo das notícias, quais são os seus graus de padronização e o uso potencial dos metadados coletados para avaliar o modelo desenvolvido;
4. Avaliar a ontologia em contexto de uso controlado.

1.2. HIPÓTESE DE TRABALHO

Como hipótese de trabalho expor se, utilizando como fonte de aquisição de conhecimento um conjunto de informação e uma combinação de indexação humana e indexação semiautomática, é possível modelar uma ontologia do domínio política-economia para a representação de ações desse domínio. Para atender essa hipótese, é desenvolvida uma metodologia baseada principalmente sobre os princípios metodológicos de Mendonça y Almeida (2014) que compreende: a) uma revisão teórica sobre a web semântica, ontologias e indexação; b) a criação de uma amostra de notícias de jornais; c) a definição de um marco teórico para a modelagem de ações com base na Gramática de Casos de Fillmore (1967) e García Gutiérrez (2014); d) a construção e a implementação de *Ontología de Acciones (ONA)*; e) a construção de uma base de conhecimento através da indexação humana e semiautomática de notícias da amostra orientada à modelagem e implementação da *Ontología de Acciones de Política e Economía (ONAPE)*; f) a análise de vocabulários já existentes, principalmente em *Linked Open Data (LOD)* para o mapeamento inicial destes vocabulários com ONA; e g) a avaliação baseada em dados das ontologias desenvolvidas.

2. FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA

2.1. AS ONTOLOGIAS

2.1.1. CONCEITO DE ONTOLOGIA

A ontologia, da qual se pode destacar como característica distintiva a riqueza das relações entre os seus termos (Stuart, 2016, p. 9), é definida por Guarino (1998) como um vocabulário específico utilizado para descrever uma realidade, além de um conjunto de pressupostos explícitos sobre o significado pretendido das palavras do vocabulário. Daqui resulta que uma ontologia:

1. é composta por um conjunto de termos (o vocabulário),
2. dotado com certos significados determinados de algum modo (pressupostos),
3. que permite descrever uma realidade.

Como os tesauros, as ontologias são úteis para extrair e recuperar conhecimento específico (Kallipolitis et al., 2012), mas são mais complexas e seus objetivos e possibilidades vão muito além disto. As ontologias são semelhantes aos vocabulários controlados desenvolvidos para serem utilizados em uma disciplina, em uma comunidade de prática ou em um domínio concreto, mas diferem na forma como elas são construídas ou representadas (Abbas, 2010, p. 163) e o seu potencial uso. Com uma ontologia, uma pessoa não está apenas relacionada a um evento, também pode estar presente neste, organizá-lo, participar dele, ser uma autoridade dele ou até instigá-lo (Stuart, 2016, p. 9).

Nesta tese a ontologia é considerada como o último produto na evolução da representação do conhecimento (Codina y Pedraza-Jiménez, 2015), com semelhanças óbvias para outros vocabulários de representação ou organização do conhecimento. Isto é consistente com autores como Codina e Pedraza-Jiménez (2017), Stuart (2016), Codina e Pedraza-Jiménez (2011), Currás (2010), Garcia-Torres (2008), Gil Leiva (2008) e García Jiménez (2004).

2.1.2. ELEMENTOS DA ONTOLOGIA

Uma ontologia é composta de classes, propriedades, relações, indivíduos, restrições e axiomas. A este respeito, segue a descrição destes elementos, baseada em Horridge (2011), bem como Allemang e Hendler (2011, pp. 153-185), Gil Leiva (2008, pp. 227-228), Gómez, Moreno, Pazos e Sierra-Alonso (2000), Gruber (1995) e Ochoa Hernández (2011).

1. **Classes:** são conjuntos que contêm indivíduos. Trata-se de representações concretas de conceitos, que se descrevem por meio de declarações. Estas declarações especificam os requisitos para serem membros das classes. Protégé indica os seguintes tipos de classes:

- **Classes disjuntas:** àquelas de que um indivíduo não pode ser instância ao mesmo tempo. Isto significa que, diante de duas classes disjuntas, um indivíduo é instância de uma e não pode ser instância da outra
- **Classes complementares:** quando é definido que a classe A é complementar da classe B, um indivíduo da primeira classe não pode ser simultaneamente membro da classe B.

Há outros tipos de classes, que são criadas por meio da definição de restrições. Estas descrevem o conjunto de indivíduos da classe com base nas relações em que participam:

- **Classes primitivas:** os seus indivíduos precisam satisfazer certas condições necessárias para ser membros destes tipos de classe.
 - **Classes equivalentes:** o simples fato de cumprir estas condições não obriga que um indivíduo seja membro de uma classe. Para isto, as condições devem se estabelecer como suficientes. Para Horridge (2011, p. 57) uma classe que no mínimo tem condições necessárias e suficientes é denominada **classe definida**.
 - **Classes enumeradas:** são criadas pela definição prévia dos seus indivíduos. Por exemplo, *DiasDaSemana* tem 7 instâncias: {*SegundaFeira*, *TerçaFeira*, *QuartaFeira*, *QuintaFeira*, *SextaFeira*, *Sabado*, *Domingo*}.
2. **Propriedades:** também chamadas de *slots* ou atributos, exprimem as relações entre os indivíduos ou entre estes elementos e os dados. Também podem ser estabelecidas subpropriedades. As propriedades têm **domínio** e **range**. O domínio de uma propriedade inclui o conjunto de indivíduos que está relacionado com o conjunto de indivíduos do seu range (ver Figura 76).

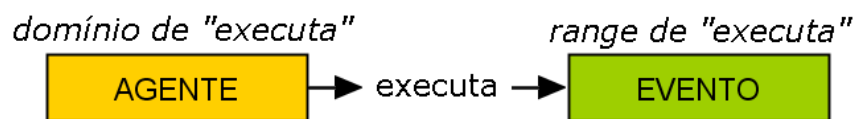


Figura 76: Domínio e range da propriedade "executa"

Existem três tipos de propriedades:

- **Propriedades de anotação:** oferecem informação a respeito das classes, indivíduos e propriedades objeto ou dado. Nas ontologias desenvolvidas por conta desta tese são utilizadas as propriedades *comment*, para as descrições de elementos e *isDefinedBy*, para indicar qual é a fonte da descrição.

- **Propriedades tipo dado:** conectam indivíduos e valores (dados) por meio de um RDF literal ou um esquema de metadados XML. Para o estabelecimento destas propriedades são empregadas as restrições *hasValue*.
- **Propriedades tipo objeto:** também denominadas de relações, existem entre classes de uma ontologia e podem ser de múltiplos tipos, por exemplo, *temCausa* ou *dependeDe*.

Algumas dessas relações são semelhantes às relações de um tesouro, sejam hierárquicas, equivalentes ou subtipos de relações associativas.

O conjunto de relações de uma ontologia é denominado de conjunto relacional básico.

3. **Indivíduos** ou **instâncias:** representam objetos concretos de um domínio e estão relacionados por meio de uma relação enumerativa com uma classe.
4. As **restrições** determinam quais indivíduos formam parte de uma classe dependendo das relações em que estes participam. Existem três tipos: de quantificação, de cardinalidade e *hasValue*.
5. Os **axiomas** são expressões que sempre são certas e se podem programar de várias formas.

Estes elementos podem variar significativamente de uma ontologia para outra. Além disso existe uma distinção entre a estrutura da ontologia (conjunto de elementos) e as suas instâncias ou indivíduos. A soma destes componentes aporta o conjunto de dados de uma ontologia ou base de conhecimento (Stuart, 2016, pp. 12-21).

Por outro lado, existem várias formas de denotar classes e propriedades, pois não pode haver espaços nem caracteres especiais (Nicolino e Ferneda, 2014). O estilo empregado para ONA e ONAPE é a notação *CamelBack*, que conta com duas convenções opcionais: os nomes das classes são iniciados por letra maiúscula e também a próxima palavra; os nomes das propriedades começam por minúscula embora a próxima palavra por letra maiúscula. Por exemplo, *Queijo eIngredienteDa PizzaTopping* (Horridge, 2011, pp. 17, 21).

2.1.3. METODOLOGIAS PARA CONSTRUIR UMA ONTOLOGIA

Existem múltiplas opções para a modelagem de uma ontologia. Neste sentido, García-Torres et al. (2008) identificam quatro formas: começar do zero, por meio de reengenharia de ontologias (reutilizando ontologias já existentes) e métodos baseados nas técnicas de aprendizagem automática através da fusão de ontologias. Na realidade, estas formas podem

se combinar, pois não existe uma metodologia definitiva, embora sejam identificadas várias propostas com ampla utilização (Silva, 2008, pp. 120-121). Alguns autores, por exemplo, Benjamins e Gómez-Pérez (2000), falam de um processo artesanal mais do que uma engenharia, onde a análise do domínio e avaliação da ontologia constituem as duas etapas que precisam de mais esforço.

Na literatura pode-se encontrar várias publicações que estudam as diferentes metodologias. Uma das análises mais aprofundadas é a de Silva (2008, pp. 218-263), seguida de Mendonça e Almeida (2014). No primeiro caso, a autora analisa até nove metodologias e cria a sua própria; no segundo caso, os autores definem os princípios necessários para a modelagem de uma ontologia, levando em conta as metodologias prévias. As duas referências são empregadas nesta pesquisa para a modelagem de ONA. Pode-se encontrar mais informação na Tabela 5, do capítulo 2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA desta memória.

2.1.4. AVALIAÇÃO DA ONTOLOGIA

A avaliação de uma ontologia compreende duas partes: verificação e validação (Asunción Gómez-Pérez, 2004; Vrandečić, 2009, p. 295).

A **verificação** tem a ver a correção da ontologia, isto é, se a ontologia satisfaz os requisitos definidos durante a sua modelagem. Editores de ontologias como Protégé incluem *reasoners* que assistem esta avaliação, por exemplo Hermit.

Já a **validação** analisa se a ontologia é correta, dito de outro modo, se realmente modela o domínio para o qual foi desenhada. Deve ser levado em conta que a modelagem de ontologias pode estar influenciada pelo seu autor(a) ou ter certa dose de subjetividade por causa da sua experiência, bem como dos seus conhecimentos. Por outro lado, o contexto de modelagem, desde o ponto de vista temporal, espacial e cultural pode ser alterado, pois o próprio domínio evolui. Assim, é importante avaliar a adequação de toda ferramenta de representação do conhecimento ao seu domínio, estudar se é útil para essa função de representação dos documentos. Neste sentido, Hlomani (2014) desenvolve um marco para a avaliação de ontologias baseado em dados, comparando uma ontologia de domínio com dados desse domínio.

Entre as medidas de validação que o autor destaca, encontram-se a precisão e a exaustividade, que tem a ver com a exatidão da ontologia. Estas medidas são empregadas nesta pesquisa para avaliar ONAPE, além da medida-F, uma combinação delas.

2.1.5. REUTILIZAÇÃO DE ONTOLOGIAS: MAPEAMENTO

A reutilização de ontologias é possível graças ao mapeamento entre os vocabulários. O mapeamento pode ser definido como uma forma de vincular os conceitos de duas

ontologias, visando obter um conjunto de relações semânticas entre elementos destas ferramentas (Shvaiko e Euzenat, 2005). Este conjunto é denominado de alinhamento, embora alguns autores definam mapeamento e alinhamento de vocabulários como termos sinônimos, como Feria et al. (2013). Assim é considerado nesta pesquisa.

Scharffe et al. (2014) desenvolveram uma biblioteca de padrões comuns de forma a auxiliar a busca por correspondências entre os elementos de dois vocabulários. Cada padrão é uma solução para um problema de alinhamento, que pode ser definido ao nível de classe, propriedade ou atributo. Os tipos de padrões podem ser de equivalência exata ou parcial, incompatibilidade, sobreposição, de elemento mais geral e elemento mais específico (Shvaiko y Euzenat, 2005). A Tabela 7, do capítulo 2. FUNDAMENTACIÓN TEÓRICO-METODOLÓGICA, mostra os tipos de mapeamento e padrões que serão utilizados nesta pesquisa como ponto de partida para a etapa de mapeamento de ONA e ONAPE com outros vocabulários.

2.2. A INDEXAÇÃO

A indexação é um processo que visa determinar o assunto de um documento, exprimindo-o por meio de termos chamados de indexação (Nicolino e Ferneda, 2014; Salton e McGill, 1983, p. 53), que descrevem o seu conteúdo para a recuperação de um documento em um sistema. A indexação também é aplicada nas equações de busca (perguntas dos usuários de um sistema por causa de uma necessidade informativa).

Frequentemente, a indexação é descrita como um processo em duas etapas: análise do documento para determinar o seu assunto e tradução deste assunto para um conjunto de termos. O resultado é uma representação do conteúdo de um documento utilizada para questionar o sistema documental e acessar um conjunto de documentos relacionados. Por esta razão, Ingersoll, Morton e Farris (2013), definem indexação como um processo para que um ou mais documentos se tornem pesquisáveis por um motor de busca.

O resultado pode ser alterado ao considerar o tipo do documento e o tipo de aproximação da indexação: 1) orientada ao usuário, onde são valorizados o documento e as necessidades informativas dos usuários; 2) orientada ao documento, que somente considera o documento e os seus autores. Ainda existem duas variações das anteriores; 3) dependendo do domínio, baseado nas necessidades dos usuários e os objetivos, propósitos e atividades no domínio do documento; 4) uma última aproximação que analisa o documento e as perguntas que o(a) indexador(a) acredita que podem ser formuladas na interrogação do sistema. Nesta tese, na fase de indexação humana é utilizada esta última perspectiva.

2.2.1. INDEXAÇÃO HUMANA E INDEXAÇÃO AUTOMÁTICA

A indexação humana é o processo cognitivo (Anderson e Pérez-Carballo, 2001b) em que uma pessoa define que conceitos representam o conteúdo de um documento ou de uma necessidade de informação, exprimida em uma equação de busca. A indexação automática, também denominada de indexação inteligente (Lamarca Lapuente, 2013a), utiliza técnicas estatísticas e/ou linguísticas para atribuir os conceitos aos documentos.

Uma evolução da indexação automática (Lamarca Lapuente, 2013a) e uma das mais áreas mais pesquisadas é o processamento da linguagem natural (PLN). Este é definido por Ruíz-Martínez (2011, p. 41) como “a tentativa de simular o comportamento linguístico humano, para que o sistema de símbolos que constitui a língua seja adquirido e processado por um computador, que reconhece, compreende, interpreta e gera linguagem humana”. O PLN pode ser usado para a extração de termos de documentos em linguagem natural, para a incorporação de elementos em uma ontologia ou um tesouro, ou bem para marcar semanticamente textos por meio de termos de um vocabulário (Stuart, 2016, p. 23). A indexação humana coincidiria com duas etapas do PLN: o processamento semântico e o processamento pragmático. Considerando o conteúdo e o propósito do PLN é mais preciso descrever este mais como semiautomático do que automático (Stuart, 2016, p. 23).

É inegável que o futuro da indexação tem a ver com a indexação automática, pois o crescimento da informação disponibilizada na Internet não permite a manutenção de processos puramente intelectuais (Lamarca Lapuente, 2013a). No âmbito jornalístico, este aumento da informação é tão grande que os tempos de produção e difusão de informação caíram, sendo aceitos “métodos automáticos de análise e visualização de dados que auxiliam na criação de conhecimento em um tempo crítico” (Pellegrini, 2012). Deste modo, parte do PLN está sendo orientado à descrição e à recuperação de informação, sem substituir o raciocínio humano, mas assistindo-o. Enquanto estes sistemas são aperfeiçoados a combinação de técnicas com maior ou menor nível de automação e a revisão humana podem ser ótimas soluções.

2.2.2. EXTRAÇÃO DE TERMOS COM N-GRAMAS

Uma das técnicas de indexação automática é o n-gramas, também utilizada para a recuperação de informação. Este modelo é baseado na ideia de análise de textos como sequências de elementos aos que precedem e seguem outros e como essas sequências mudam ao longo de todo o texto (ver Figura 20). Uma vantagem do n-gramas é que pode ser aplicado para qualquer conjunto de textos, independentemente do seu idioma (Martínez-Fernández et al., 2004) embora, os conjuntos de documentos devam ser tratados por idioma e de forma separada. Se, além disso, uma lista de palavras vazias é empregada, o número de elementos a serem comparados se reduz (Anderson e Pérez-Carballo, 2001a).

Entre os aplicativos que utilizam n-gramas encontra-se e-Termos (*Termos Eletrônicos*)⁵³, criado pelo *Centro Nacional de Pesquisa Tecnológica em Informática para a Agricultura - Embrapa Informática Agropecuária* (CNPTIA) em colaboração com a Universidade de São Paulo (USP) e a Universidade Federal de São Carlos (UFSCAR). Trata-se de um ambiente computacional colaborativo web de acesso livre dedicado à gestão terminológica (Oliveira, 2009a) para a construção de vocabulários de descrição de documentos. O e-Termos funciona por etapas, permitindo a aplicação da técnica n-gramas sobre conjuntos de documentos, de 1 grama (um termo) até 7 (sete termos). Desta forma, a co-ocorrência de dois ou mais termos também é levada em conta nesta pesquisa, pois este aplicativo é utilizado para a indexação semiautomática de textos.

2.2.3. INDEXAÇÃO DE NOTÍCIAS. A GRAMÁTICA DE CASOS

Uma aproximação de origem linguística da indexação de notícias jornalísticas é a Gramática de Casos de Fillmore (1967). Consiste na análise de uma oração de acordo com as relações sintático-semânticas entre os “casos” universais que a compõem. Os casos são “seres, coisas, fatos y situações extralinguísticas comuns à experiência humana” (García-Hernández, 1986) que facilitam a representação do conteúdo de um documento qualquer. Casos típicos para os substantivos são agente, objeto, instrumento e objeto indireto (Salton e McGill, 1983, p. 263). Nesta linha, García Gutiérrez (2014) realiza um “inventário exaustivo” de perguntas necessárias para a obtenção de dados factuais de uma notícia e propõe que estas questões sejam respondidas através da Gramática de Casos.

Ws submetidas por García Gutiérrez	Respostas a Ws
<i>Quem/Que?</i>	Sujeito que precede à <i>ação</i>
<i>A/Contra quem?</i>	Receptor/es direito/s da <i>ação</i>
<i>Que coisa é objeto da ação</i>	Objeto beneficiado ou prejudicado pela <i>ação</i>
<i>Por meio de que e que se realiza a ação?</i>	Processos não tangíveis especializados ou ordinários
<i>Através de que se realiza a ação?</i>	Instrumental. Objeto tangível através do qual se executa a <i>ação</i> . Precisa da mediação ou manipulação por sujeitos
<i>Porque é realizada a ação?</i>	Causal. Extensível aos motivos e antecedentes
<i>Para que é realizada a ação?</i>	Intenções e objetivos não alcançados
<i>Quais são as consequências da ação?</i>	Resultados alcançados e constatados
<i>Quando é que se realiza a ação?</i>	Data, tempo, período
<i>Onde é que se realiza a ação?</i>	Lugar. Espaços genéricos ou específicos
<i>Com quem o que é que se realiza a ação?</i>	Associativo. Sujeito colaborador, não protagonista, acompanhante
<i>No que estado é que se e realiza a ação?</i>	Aparece junto ao elemento que matiza

Tabela 1: 12W de uma notícia, segundo García Gutiérrez

53 <https://www.etermos.cnptia.embrapa.br/index.php> (último acesso: 21/06/2017)

Para a definição das 12Ws, García-Gutiérrez, baseando-se principalmente em Fillmore (1967) construiu macroproposições, para aplicar às perguntas da Tabela 1. Estas questões são utilizadas na metodologia desta pesquisa para a modelagem de ONA e, depois, de ONAPE.

2.2.4. AVALIAÇÃO DA INDEXAÇÃO

Na avaliação dos resultados de uma indexação destacam diferentes características, as quais têm sido analisadas por vários autores, entre os que podem-se sublinhar Cleveland e Cleveland (2013, pp. 158-160), Gil Leiva (2012, pp. 78-84), Chowdhury (2010, pp. 99-100), Martínez Méndez (2004), Martínez Méndez (2002), Salton e McGill (1983) e Cavalcanti (1978, pp. 56-57). Estas características permitem determinar o grau em que um conjunto de termos atribuídos aos documentos representam ou não o seu conteúdo, mas também se esses termos respondem adequadamente à pergunta formulada por um usuário.

Entre essas características, destacam-se duas medidas muito utilizadas na avaliação da indexação, a exaustividade e a precisão na recuperação da informação. Elas estão relacionadas com a quantidade de conceitos atribuídos a um documento e a equação de busca para a sua recuperação em um sistema. Em uma procura de informação, a exaustividade leva em conta quantos documentos relevantes foram recuperados em relação ao conjunto total de documentos relevantes existentes no sistema (ver Figura 24). A precisão valora a quantidade de documentos relevantes recuperados em relação ao total de documentos recuperados (Salton e McGill, 1983, pp. 55, 164) (ver Figura 25). Isso pode ser resumido da seguinte forma: “um sistema ideal dará ao usuário(a) todos os documentos que ele(a) precise [exaustividade], e não mais [precisão]” (Cleveland e Cleveland, 2013, p. 159). Estas duas medidas são inversas: maior exaustividade, menor precisão, e vice-versa.

A exaustividade de uma recuperação obterá melhores resultados por méio da utilização de termos mais genéricos. Por outro lado, a precisão é recompensada pelo uso de termos específicos ou pela combinação de vários deles (Salton, 1975, pp. 92, 100). Em outras palavras, mais exaustividade na indexação, maior exaustividade na recuperação (sem ruído informativo); quanto maior a especificidade da indexação, melhor será a precisão da recuperação de informação (menor ruído informativo) (Salton e McGill, 1983, p. 162). Ambas as medidas estão frequentemente relacionadas com a medida o valor F, também denominada de medida harmônica (Figura 27).

Estas medidas são aplicadas na avaliação das ontologias modeladas para analisar a sua capacidade para representar o conteúdo de uma amostra de notícias.

3. PROCEDIMENTOS METODOLÓGICOS E RESULTADOS

As metodologias desenvolvidas para a modelagem e construção da *ONtología de Acciones de Política y Economía (ONAPE)* são baseadas nos princípios metodológicos de Mendonça e Almeida (2014). Como resultado, são obtidas duas ontologias, ONAPE e *ONtología de Acciones (ONA)*. As etapas que seguem nem sempre têm uma linha temporal e podem sobrepor-se umas às outras.

3.1. CONSTRUÇÃO DA AMOSTRA

Para o desenho da amostra, procedeu-se com uma seleção de zonas geopolíticas, países e jornais, por meio de uma amostragem orientada. Os dados dos meios de comunicação foram extraídos da web *4 International Media and Newspapers (4IMN)*.

Depois se escolheu o jornal mais lido de cada país, levando em conta os dados de acesso pela web e compra de edição impressa de 4IMN. A Tabela 2 mostra os 21 meios de comunicação utilizados nesta pesquisa.

Zona do mundo	País	Jornal	Código
Europa	Alemanha	<i>Süddeutsche Zeitung</i>	101
	França	<i>Le Monde</i>	111
	Reino Unido	<i>The Daily Telegraph</i>	121
		<i>Financial Times</i>	122
<i>The Economist</i>		123	
América do Norte	Estados Unidos	<i>The New York Times</i>	201
<i>The Wall Street Journal</i>		202	
Latinoamérica	Brasil	<i>O Globo</i>	301
	México	<i>El Univesal</i>	311
Ex-repúblicas soviéticas	Rússia	<i>Pravda</i>	401
África Subsariana	Nigéria	<i>Nigerian Tribune</i>	501
	África do Sul	<i>Independent Online</i>	511
Ásia	China	<i>China Daily</i>	601
	Índia	<i>The Times of India</i>	611
	Japão	<i>The Asahi Shimbun</i>	621
Leste	Arábia Saudita	<i>Arab News</i>	701
	Emiratos Árabes Unidos	<i>Gulf News</i>	711
	Israel	<i>Yedioth Aharonot</i>	721
	Turquia	<i>Today's Zaman</i>	731
Oceania	Austrália	<i>The Australian Financial Review</i>	801
Norte de África	Egipto	<i>The Daily News Egypt</i>	901

Tabela 2: Jornais selecionados

Diariamente para cada jornal, de julho à dezembro de 2012 e 2013, uma notícia por jornal foi selecionada até construir uma amostra de 5.582 notícias. Esta foi dividida em duas partes, uma para a modelagem de ONAPE e outra para a avaliação da ontologia: a amostra₁ inclui artigos de julho à outubro de 2012 (1.837 notícias). Este conjunto é utilizado para os processos de indexação humana e automática, permitindo a modificação de ONA e a modelagem de ONAPE; a amostra₂ inclui peças agrupadas pela sua temática, sendo no total cinco temas aleatórios e 23 notícias (ver Tabela 3).

Palava-chave	Tema
Ilhas Senkaku (<i>Senkaku Islands</i>)	Conflito territorial das Ilhas Senkaku
Mandela	Homenagens a Nelson Mandela após sua morte em 2013
Inmigración (<i>immigration</i>)	Restrições à entrada de imigrantes no Reino Unido e nos Estados Unidos
Corrupção (<i>corruption</i>)	Caso de corrupção em Turquia
Energia (<i>energy</i>)	Mercado energético

Tabela 3: Temáticas das notícias da subamostra₂

3.2. MODELAGEM DA ONTOLOGIA ONA BASEADA NA GRAMÁTICA DE CASOS

Partindo do marco teórico da Gramática de Casos de Fillmore (1967) e García Gutiérrez (2014) é modelada a **ONTologia de Acciones (ONA)** para a descrição geral de ações. Modelar é o processo pelo qual o conhecimento é organizado para o seu uso por parte de uma comunidade, segundo Allemang e Hendler (2011). Desta maneira, primeiro são analisados os componentes desta teoria, estabelecendo-se relações entre estes.

Cada caso é classificado em uma ou mais entidades básicas. “Uma entidade é uma forma particular de categorização de tipo evento, lugar, organização, pessoa, produto e congêneres” (Rocha et al., 2009). Cada entidade básica e caso constitui uma classe potencial de ONA:

1. **Ação:** fato produzido em um espaço temporal e um espaço físico, no qual participa pelo menos um sujeito. Uma ação pode ter: **causas**, motivos que dão origem à ação; **consequências**, resultam da execução da ação.
2. **Sujeito:** “ser do qual é dito ou anunciado algo” (RAE, 2014k). São sujeitos as pessoas, os coletivos e as organizações. Um sujeito pode agir como:
 - **Agente** que executa uma ação;
 - **Receptor**, sujeito que recebe os efeitos de uma ação;
 - **Acompanhante**, sujeito que acompanha a quem executa uma ação.

3. **Objeto:** “tudo o que tem entidade, seja corporal ou espiritual, natural ou artificial, real ou abstrata” (RAE, 2014b). Um objeto pode ter estas funções:
- **Agente** que executa uma ação;
 - **Objeto afetado** pela ação executada;
 - **Produto**, resultado da execução de uma ação;
 - **Instrumento**, de caráter tangível, “coisa ou pessoa de quem alguém se serve para fazer algo ou conseguir um fim” (RAE, 2014g);
 - **Estado**, estado involuntário no qual se pode encontrar quem executa ou sofre uma ação;
 - **Modo**, se faz referência ao procedimento, técnica, maneira ou gênero com que se executa uma ação (García Gutiérrez, 2014);
 - **Finalidade**, se trata de um desejo, objetivo ou propósito não alcançado pelo que executar uma ação; intenção de quem executa uma ação;
1. **Espaço físico:** espaço ocupado por um corpo qualquer. Espaço no qual acontece uma ação. Esta definição é baseada em RAE (2014j).
4. **Espaço temporal:** “tempo em que acontece ou é feito algo” (RAE, 2014e), momento.

Entre estas classes existem relações que as vinculam: *acompanhaA*; *afetaA*; *executa*; *eAcompanhadoPor*; *eAfetadoPor*; *eCausaDe*; *eConsequenciaDe*; *eExecutadaPor*; *eEspaçoFísicoDe*; *eEspaçoTemporalDe*; *eSituacaoDe*; *eFinalidadeDe*; *eModoDe*; *eProdutoDe*; *einstrumentoUsadoEm*; *temCausa*; *temConsequencia*; *temEspaçoFísico*; *temEspaçoTemporal*; *temSituacao*; *temFinalidade*; *temInstrumento*; *temModo*; *temProduto*. Desta forma, se estabelece o domínio e o range de cada propriedade. São descritas na Tabela 6. Quanto à formalização de classes e relações se adota a notação *CamelBack*, e as recomendações de Aguado-de-Cea et al. (2015).

A Gramática de Casos é definida no nível de linguagem natural. No entanto, as ontologias são definidas no nível formal. Por esta questão, o modelo levantado inicialmente e analisado e modificado com vistas a sua formalização em OWL. Como exemplo desta análise se apresentam as relações entre as classes *Instrumento*, *Agente* e *Ação*. Tal como foram definidas pela Gramática de Casos não é possível identificar corretamente que função tem cada elemento em uma *Ação*, não seria possível atribuir a participação de um *Instrumento* em uma *Ação* específica por determinado *Agente*. Propõem-se três soluções:

- a) Modificar a modelagem relacionando *Instrumento* e *Ação*. No caso de que participaram dos *Agentes*, quem é que usa o *Instrumento*?
- b) Mudar a modelagem mantendo a vinculação original, *Agente-Instrumento* e *Instrumento-Ação*, não manteria a independência entre um *Agente* e um *Instrumento* se o *Agente* participa em mais de uma *Ação*. Se este tem sido relacionado com dois *Instrumentos*, em duas *Ações*, com que *Instrumento* participou em cada?
- c) Modificar a modelagem incorporando um nó anônimo ou classe auxiliar, quem tem relações com os três elementos indicados. De um lado, se exprimiria a participação de um *Agente* numa *Ação* com um *Instrumento* específico; de outro, assegura-se a independência do *Agente* para qualquer outro tipo de *Ação*. Este tipo de classe auxiliar executaria a *Ação*.

Esta solução é aplicada para outras classes do modelo, também afetadas pelo mesmo assunto que o *Instrumento*: *Modo*, *Finalidade* e *Estado*. Desta maneira, uma classe anônima, denominada *Sujeito*, é que executa a *Ação* e relaciona *Agente* com: o *Estado* em que se encontra, o *Instrumento* que utiliza, o *Modo* em que realiza a *Ação* e a *Finalidade* com que a realiza. Do mesmo modo, a classe *Objeto* é um elemento auxiliar que relaciona o *Receptor* de uma *Ação* e o *Estado* em que se encontra quando recebe os efeitos da mesma. Para facilitar a interpretação da modelagem, é que se inverteram os nomes de *Sujeito* e *Agente*.

A ontologia ONA tem portanto relações do tipo *Sujeito e SujeitoDe Ação* e *Objeto e ObjetoDe Ação*, criando uma estrutura similar às triplas RDF (sujeito-predicado-objeto). Por outro lado, foram suprimidas as classes *Finalidade*, *Estado*, *Instrumento*, *Acompanhante*, *Causa* e *Produto*, ficando esses conceitos expressos diante relações. Com isto a modelagem é simplificada. As equivalências entre casos e elementos de ONA se encontram na tabela seguinte.

Casos	Elementos do modelo
Ação	<i>Ação</i>
Objetivo	<i>Objeto (Coisa)</i>
Factitivo	<i>eResultadoDe, temResultado</i>
Instrumental	<i>eInstrumentoDe, temInstrumento</i>
Estado	<i>eSituacaoDe, temSituacao</i>
Agentivo	<i>Sujeito (Agente ou Coisa)</i>
Dativo	<i>Objeto (Agente)</i>
Asociativo	<i>eAcompanhanteDe, temAcompanhante</i>
Modo	<i>eModoDe, temModo</i>
Locativo	<i>EspaçoTemporal eEspaçoTemporalDe, temEspaçoTemporal</i>
Data	<i>EspaçoFísico eEspaçoFísicoDe, temEspaçoFísico</i>
Causativo	<i>eCausaDe, temCausa</i>
Finalidade	<i>eFinalidadeDe, temFinalidade</i>

Tabela 4: Equivalências entre Casos e elementos de ONA

Para evitar problemas de homonímia do termo *Estado* como *situação* e *Estado* como “conjunto de instituições que possuem a autoridade e poder para estabelecer as normas que regulam uma sociedade, tendo soberania interna e externa sobre um País”, se utiliza o termo *Situação* para o primeiro caso. A Tabela 5 mostra as classes de ONA.

Classe	Descrição
Ação	Entidade básica que exprime o fato produzido em um <i>EspaçoTemporal</i> e um <i>EspaçoFísico</i> , onde participa no mínimo um <i>Sujeito</i>
Agente	Ser que pode participar em uma <i>Ação</i> , executando-a, acompanhado a quem a executa ou recebendo os seus efeitos
Coisa	Tudo o que tem entidade, tangível ou não
EspaçoFísico	Espaço físico o lugar em que acontece uma <i>Ação</i>
EspaçoTemporal	Espaço temporal (data, período, tempo) em que acontece uma <i>Ação</i>
Objeto	Classe auxiliar que vincula um <i>Agente</i> ou <i>Coisa</i> e a <i>Ação</i> da qual recebe os efeitos
Sujeito	Classe auxiliar que vincula um <i>Agente</i> ou <i>Coisa</i> e a <i>Ação</i> que executa

Tabela 5: Classes de ONA

RESUMO EM PORTUGUÊS

As relações entre classes mostram-se na Tabela 6, que contém nome, descrição, domínio e range de cada relação.

Propriedade	Descrição	Domínio	Range
<i>eAcompanhanteDe</i>	Relaciona <i>Agente</i> ou <i>Coisa</i> que acompanha a quem executa uma ação por meio do <i>Sujeito</i> . Inversa: <i>temAcompanhante</i>	<i>Agente</i> <i>Coisa</i>	<i>Sujeito</i>
<i>eCausaDe</i>	Relaciona uma <i>Ação</i> que <i>eCausaDe</i> outra e essa <i>Ação</i> . Inversa: <i>temCausa</i>	<i>Ação</i>	<i>Ação</i>
<i>eConsequenciaDe</i>	Relaciona a <i>Ação</i> consequência de uma <i>Ação</i> e esta. Inversa: <i>temConsequencia</i>	<i>Ação</i>	<i>Ação</i>
<i>eExecutanteDe</i>	Relaciona o <i>Agente</i> ou <i>Cosa</i> que executa uma <i>Ação</i> e o <i>Sujeito</i> através do que a executa. Inversa: <i>temExecutante</i>	<i>Agente</i> <i>Coisa</i>	<i>Sujeito</i>
<i>eEspaçoFisicoDe</i>	Relaciona o <i>EspaçoFisico</i> no que se produz uma <i>Ação</i> e esta. Inversa: <i>temEspaçoFisico</i>	<i>EspaçoFisico</i>	<i>Ação</i>
<i>eEspaçoTemporalDe</i>	Relaciona o <i>EspaçoTemporal</i> no que se produz uma <i>Ação</i> e dita <i>Ação</i> . Inversa: <i>temEspaçoTemporal</i>	<i>EspaçoTemporal</i>	<i>Ação</i>
<i>eFinalidadeDe</i>	Relaciona a intenção de um <i>Agente</i> na execução de uma <i>Ação</i> por meio de um <i>Sujeito</i> e esse <i>Sujeito</i> . Inversa: <i>temFinalidade</i>	<i>Coisa</i>	<i>Sujeito</i>
<i>eInstrumentoDe</i>	Relaciona um <i>Instrumento</i> e um <i>Agente</i> que o usa para executar uma <i>Ação</i> , por meio da classe <i>Sujeito</i> . Inversa: <i>temInstrumento</i>	<i>Coisa</i>	<i>Sujeito</i>
<i>eModoDe</i>	Relaciona uma <i>Coisa</i> utilizada como Instrumento para a execução de uma <i>Ação</i> e o <i>Sujeito</i> através do qual é utilizada. Inversa: <i>temModo</i>	<i>Coisa</i>	<i>Sujeito</i>
<i>eObjetoDe</i>	Relaciona o <i>Objeto</i> através do que um <i>Agente</i> ou <i>Coisa</i> recebe os efeitos de uma <i>Ação</i> e esta. Inversa: <i>temObjeto</i>	<i>Objeto</i>	<i>Ação</i>
<i>eReceptorDe</i>	Relaciona o <i>Agente</i> ou <i>Coisa</i> que recebe os efeitos de uma <i>Ação</i> e <i>Objeto</i> , classe auxiliar através da qual os recebe. Inversa: <i>temReceptor</i>	<i>Agente</i> <i>Coisa</i>	<i>Objeto</i>
<i>eResultadoDe</i>	Relaciona a <i>Coisa</i> produto da execução de uma <i>Ação</i> e esta. Inversa: <i>temResultado</i>	<i>Coisa</i>	<i>Ação</i>
<i>eSituacaoDe</i>	Relaciona a situação de um <i>Agente</i> ou <i>Coisa</i> , através de um <i>Sujeito</i> ou <i>Objeto</i> , na execução ou recepção de uma <i>Ação</i> . Inversa: <i>temSituacao</i>	<i>Coisa</i>	<i>Sujeito</i> <i>Objeto</i>
<i>eSujeitoDe</i>	Relaciona <i>Sujeito</i> e a <i>Ação</i> . Inversa: <i>temSujeito</i>	<i>Sujeito</i>	<i>Ação</i>
<i>temAcompanhante</i>	Relaciona um <i>Sujeito</i> e o <i>Agente</i> ou <i>Coisa</i> que acompanha a outro na execução de uma <i>Ação</i> . Inversa: <i>eAcompanhanteDe</i>	<i>Sujeito</i>	<i>Agente</i> <i>Coisa</i>
<i>temCausa</i>	Propriedade que relaciona uma <i>Ação</i> e a <i>Causa</i> por a qual ocorre. Inversa: <i>eCausaDe</i>	<i>Ação</i>	<i>Ação</i>
<i>temConsequencia</i>	Relaciona uma <i>Ação</i> e a <i>Consequência</i> que produz. Inversa: <i>eConsequenciaDe</i>	<i>Ação</i>	<i>Ação</i>

Propriedade	Descrição	Domínio	Range
<i>temExecutante</i>	Relaciona um <i>Sujeito</i> e o <i>Agente</i> ou <i>Coisa</i> que executa uma <i>Ação</i> . Inversa: <i>eExecutanteDe</i>	<i>Sujeito</i>	<i>Agente</i> <i>Coisa</i>
<i>temEspaçoFísico</i>	Relaciona uma <i>Ação</i> e o <i>EspaçoFísico</i> no que se produz. Inversa: <i>eEspaçoFísicoDe</i>	<i>Ação</i>	<i>EspaçoFísico</i>
<i>temEspaçoTemporal</i>	Relaciona uma <i>Ação</i> e o <i>EspaçoTemporal</i> no que se produz. Inversa: <i>eEspaçoTemporalDe</i>	<i>Ação</i>	<i>EspaçoTemporal</i>
<i>temFinalidade</i>	Relaciona um <i>Sujeito</i> e a finalidade, propósito ou intenção de um <i>Agente</i> pela que executa uma <i>Ação</i> . Inversa: <i>eFinalidadeDe</i>	<i>Sujeito</i>	<i>Coisa</i>
<i>temInstrumento</i>	Relaciona um <i>Sujeito</i> e o <i>Instrumento</i> que usa para executar uma <i>Ação</i> . Inversa: <i>eInstrumentoDe</i>	<i>Sujeito</i>	<i>Coisa</i>
<i>temModo</i>	Relaciona um <i>Sujeito</i> e o <i>Modo</i> em que executa uma <i>Ação</i> . Inversa: <i>eModoDe</i>	<i>Sujeito</i>	<i>Coisa</i>
<i>temObjeto</i>	Relaciona <i>Ação</i> e o <i>Objeto</i> por meio do qual um <i>Agente</i> ou <i>Coisa</i> recebe os efeitos da execução de essa <i>Ação</i> . Inversa: <i>eObjetoDe</i>	<i>Ação</i>	<i>Objeto</i>
<i>temReceptor</i>	Relaciona <i>Objeto</i> com o <i>Agente</i> ou <i>Coisa</i> que recebe o efeito de <i>Ação</i> . Inversa: <i>eReceptorDe</i>	<i>Objeto</i>	<i>Agente</i> <i>Coisa</i>
<i>temResultado</i>	Relaciona uma <i>Ação</i> e <i>Coisa</i> que resulta da sua execução. Inversa: <i>eResultadoDe</i>	<i>Ação</i>	<i>Coisa</i>
<i>temSituação</i>	Relaciona <i>Sujeito</i> ou <i>Objeto</i> e a sua <i>Situação</i> na execução/recepção de uma <i>Ação</i> . Inversa: <i>eSituaçãoDe</i>	<i>Sujeito</i> <i>Objeto</i>	<i>Coisa</i>
<i>temSujeito</i>	Relaciona <i>Ação</i> e o <i>Sujeito</i> através do que um <i>Agente</i> ou <i>Coisa</i> a executa. Inversa: <i>eSujeitoDe</i>	<i>Ação</i>	<i>Sujeito</i>

Tabela 6: Relações iniciais de ONA

Nesta etapa foi definida boa parte do que será ONA, porém não é um produto acabado. Depois a ontologia será novamente modificada.

3.2. DESCOBRIMENTO DE CONHECIMENTO

Levando em conta a modelagem inicial de ONA e com vistas a incorporação de subclasses e propriedades, foram indexadas notícias da amostra₁, determinando o tema da notícia e a sua tradução para um conjunto de termos. Para isto, dois processos de indexação foram utilizados, indexação humana e automática. Cada termo atribuído às notícias ou extraído por e-Termos é, potencialmente, um elemento a ser recolhido por ONA, no caso de termos gerais e por ONAPE no caso de termos especializados. No total se indexaram 1.837 itens (sub-amostra₁), atribuindo a estes diversos termos representantes de determinados conceitos no contexto da Política e da Economia. Desta forma, foram obtidas listas de termos suscetíveis de ser incorporados na ONAPE.

A extração de termos por indexação automática é assistida pelo software e-Termos (*Termos Eletrônicos*). Este aplicativo utiliza técnica n-gramas, com combinações de uma

grama (um termo) a sete gramas (termo composto por 7 palavras). Primeiro, se preparam os textos que vão ser submetidos à indexação automática, traduzindo os títulos ao espanhol. Além disso, é atribuída uma lista de palavras vazias ao processo. Depois, se especifica o tipo de documento com o qual e-Termos trabalha. Finalmente, foram obtidas listas de termos para cada grama, desde 1 grama até 7. No caso de unigramas, os termos foram submetidos à lematização com o software Grampal. Os dados estatísticos iniciais se indicam na seguinte tabela.

TIPO DE DADO ESTATÍSTICO	DADOS TOTAIS
<i>Palavras identificadas nos títulos</i>	21396
<i>Palavras diferentes</i>	5580
<i>Acrônimos e nomes próprios encontrados</i>	847

Tabela 7: Dados estatísticos iniciais obtidos em e-Termos

Tanto no caso de listas de termos obtidos por meio da indexação humana quanto da automática, cada termo e cada elemento é analisado individualmente. Desta forma, se valoriza a pertinência da incorporação do termo no vocabulário. Como em qualquer processo humano, a decisão do que é uma classe, uma relação ou outro elemento não é absoluta, diferentes pessoas podem diferir na opinião do que é correto e o que não é para a ontologia (Stuart, 2016, p. 31).

A partir dos resultados da indexação humana se considera conveniente incluir mais quatro elementos na ontologia ONA, descritos na Tabela 8.

Propriedade	Descrição	Domínio	Range
<i>eAntedentePara</i>	Propriedade que expressa a relação de antecedência de uma <i>Ação</i> sobre outra, quando essa relação não é de causalidade nem de consequência. Inversa: <i>temAntecedente</i>	<i>Ação</i>	<i>Ação</i>
<i>eParteDe</i>	Expressa relações hierárquicas de tipo partonômico (parte-todo). Inversa: <i>temParte</i>	---	---
<i>temAntecedente</i>	Permite expressar a relação entre uma <i>Ação</i> e outra que a antecede, quando essa relação não é de causalidade nem de consequência. Inversa: <i>eAntedentePara</i>	<i>Ação</i>	<i>Ação</i>
<i>temParte</i>	Propriedade definida para exprimir relações hierárquicas de tipo partonômico (todo-parte). Inversa: <i>eParteDe</i>	---	---

Tabela 8: Propriedades tipo objeto incorporadas a ONA

Estas propriedades permitem estabelecer relações entre classes quando uma é parte de outra, bem como relações de antecedência. Neste caso, existe uma ligação temporal entre duas ações, que é distinta da relação de uma ação e as suas causas e consequências.

Na indexação automática, cada termo é revisado e contrastado com: 1) as notícias que aparece; 2) com a ontologia ONAPE; 3) com as diversas fontes utilizadas para a descrição de classes e propriedades.

Contudo, grande parte dos resultados fornecidos pelo e-Termos não foram incorporados a ONAPE, especialmente no caso dos tri-gramas. Isso se deve fundamentalmente à incapacidade do aplicativo para utilizar a lista de palavras vazias, produzindo falsos positivos. A próxima tabela mostra as frequências dos termos incluídos nos vocabulários, classificados nas cinco entidades básicas de ONA e ONAPE para cada n-grama. Convém lembrar que *Objeto* e *Sujeito* são classes auxiliares, só para facilitar a relação dos distintos elementos na descrição de ações.

n-gramas	Ações	Espaços Temporais	Espaços Físicos	Coisas	Agentes	TOTAL
1	192	19	30	217	141	599
2	22	0	11	32	46	111
3	5	0	2	22	21	50
4	8	1	1	5	10	25
5	4	0	0	2	1	7
6	1	0	0	2	0	3
7	1	0	0	2	0	3
TOTAL	234	20	42	281	220	798

Tabela 9: Frequências n-gramas por classe de ONA para os termos revisados

A incorporação de um elemento fornece não somente uma classe ou propriedade, mas gera outros elementos necessários para uma descrição adequada de *Ações*, *Agentes*, *Coisas*, etc. Por exemplo: o hexa-grama e instância *Embaixada dos Estados Unidos da America em Yemen* geraria 2 classes, 3 instâncias e 3 propriedades: *eUm* (relação enumerativa entre classe e instância); *representaAoEstado* e *representaAoEstadoAnte*.

3.3. MODELAGEM DE ONAPE A PARTIR DOS RESULTADOS DA INDEXAÇÃO HUMANA

Levando em conta ONA e o contexto (notícias) em que aparecem os termos, se modela ONAPE. Isto implica a inclusão dos elementos específicos do domínio político-econômico. Por exemplo, a partir da expressão *Internet é uma das ferramentas do terrorismo*, se modela *Internet* como uma *Coisa* com função de *Ferramenta*; *eFerramentaDe* é uma relação; *Terrorismo* é uma *Ação*.

Para a modelagem de ONAPE é adotado o enfoque *middle-out*, que intercala as abordagens *top-down* e *bottom-up* e favorece o descobrimento de conceitos durante a construção da ontologia (Beydoun et al., 2011). Este procedimento consiste em definir os termos fundamentais e, posteriormente, definir termos mais abstratos e específicos. Por exemplo, *Mamífero* é uma generalização de *Cão*, e *Cocker spaniel* é uma especialização (Uschold e Grüninger, 1996, pp. 20-22). Para esses autores, esse enfoque facilita a relação entre

termos na área de forma mais precisa e reduz a duplicidade do trabalho. Desta forma, para a definição de conceitos e relações, bem como para a sua classificação nos dois vocabulários, são utilizadas várias fontes, como o *Diccionario de la Lengua Española* (DLE), Wikipedia, os tesouros EUROVOC, UNBIS e AGROVOC e vocabulários como schema.org, FOAF ou TIME. O mapeamento de ONA com alguns destas ferramentas foi iniciado nesta pesquisa.

Neste sentido, são consideradas as definições dos termos obtidos por indexação e são aplicados dois métodos comuns de criação de vocabulários controlados: indução, pelo que se agregam elementos ao vocabulário a partir da indexação de documentos (Keyser, 2012, p. 25); e dedução, pelo que se recolhem classes e propriedades obtidas das definições e relações dos termos de indexação com outros elementos das ferramentas já indicadas (EUROVOC, UNBIS, etc.). Para cada termo se analisa tanto o seu contexto (as notícias) quanto a sua descrição nestas fontes de referência, levando em conta a definição dos conceitos e das suas relações. Assim, agentes coletivos (governos, empresas, consumidores), que costumam ser considerados territórios (García Gutiérrez, 2014), são classificados como *Agentes*, por exemplo, *Governo de Portugal*.

Portanto, os conceitos que se incorporam a ONAPE podem vir de quatro tipos de fonte:

- Elementos obtidos por indexação sejam conceitos ou instâncias. Neste último caso, sua classe é identificada e incorporada ao modelo;
- Conceitos que formam parte da definição daqueles atribuídos na indexação humana;
- Intrarelacões hierárquicas, associativas e de equivalência entre termos e conceitos de vocabulários empregados para a definição de elementos de ONAPE. Por exemplo, *Carro TR Indústria automobilística*;
- Extrarelacões entre conceitos de um tesouro com outro através de relações. Por exemplo, os elementos do tesouro EUROVOC mantêm relações do tipo *Has Exact Match*, *Has Narrow Match* e *Has Related Match* com outras ferramentas.

No que diz respeito às ferramentas para as definições de conceitos, se utilizam tesouros de âmbito europeu e internacional, gerais e especializados: EUROVOC; GEMET (General Multilingual Environmental Thesaurus); AGROVOC; ECLAS e UNBIS.

Mas os tesouros classificam e relacionam conceitos, porém não costumam descrevê-los além de algumas notas de alcance, com a exceção de GEMET. Para as definições de elementos de ONAPE se recorre ao DLE, Wikipédia e, em menor proporção, ao *Diccionario Económico de Expansión* e Wolters Kluwer. Para o caso de Wikipedia, se não existem fontes suficientemente comprovadas ou o artigo é controverso, a sua versão em

inglês é utilizada. Além disso, são levados em conta vocabulários como schema.org, povoado com instâncias do mundo real, *Dolce+DNS Ultralite Ontology* (DUL) e *Political Roles Ontology* (PROLES).

Como resultado de todo o processo, se obtiveram os dados estatísticos da Tabela 10.

Elemento do vocabulário	ONA	ONAPE
Termos iniciais	663	3.982
Classes	20	1.047
Propriedades tipo objeto	32	69
Instâncias	23	83
Relações hierárquicas de classe	13	1.175
Relações hierárquicas de propriedade	24	36
Propriedades tipo dado	10	14
Propriedades de anotação	196	3.370
Tipo comentário	75	1.207
Tipo definição	40	900
Labels	75	1.172
Tipo ver também	0	91

Tabela 10: Dados estatísticos das modelagens de ONA e ONAPE

O mapeamento com outros vocabulários, descrito a seguir, também é utilizado nas modelagens dos vocabulários por meio das definições de classes e propriedades (*isDefinedBy* ou propriedade tipo definição na Tabela 10).

3.4. MAPEAMENTO COM OUTROS VOCABULÁRIOS

Diversos vocabulários foram analisados para o seu mapeamento com ONA e ONAPE. Os nomes das ferramentas da primeira seleção são mostrados abaixo:

- **Ontologias gerais:** Basic Formal Ontology (BFO), Conceptual Reference Model (CIDOC CRM), DBPEDIA, Dolce+DnS Ultralite Ontology (DUL), Dublin Core Metadata Initiative (DCMI), Schema.org e Upper Mapping and Binding Exchange Layer (UMBEL).
- **Ontologias de eventos:** The Event Ontology (EVENT), F-Model of Events e Ontology for Linking Open Descriptions of Events (LODE).
- **Ontologias de agentes:** Friend of a Friend (FOAF), The Person Core Vocabulary (PERSON), vCard Ontology, Vocabulary for Biographical Information (BIO) e Vocabulary for Poderopedia (PODER).

- **Ontologias para espaços temporais:** Time Event Ontology (TEO), The Time Ontology (TIME) e The Timeline Ontology (TL).
- **Ontologias para espaços físicos:** Cartcoordont (CART), FAO Geopolitical Ontology (GEOP), The Places Ontology e WGS84 Geo Positionning.
- **Ontologias para meios de comunicação:** BBC Core Concept Ontology, ePaper, Multimedia Semantic Syndication For Enhanced News Services Project (MESH), Neptuno, News Ontology Event Model (NOEM), The News Ontology Project, News Storyline Ontology (NSL), Pena (2012), OMNIPAPER, SENSO Knowledge Based Ontology, Simple News and Press Ontologies (SNaP Ontologies), SMARTPUSH Project, The World News Ontology (WNO) e Troncy (2008).
- **Ontologias de política ou economia:** BBC Politics Ontology (POLITICS), CORRUPTECA, ESSGLOBAL Vocabulary, POLIONTO, Political Roles Ontology (PROLES), Politics Ontology for Web Entity Retrieval (POWER).

Depois, já na segunda seleção de ontologias, adotou-se diversos critérios visando a identificação daquelas que melhor atendem aos requerimentos de ONA. Neste documento somente são indicadas as ferramentas selecionadas para o mapeamento com ONA, sendo prevista uma futura pesquisa para o mapeamento com ONAPE:

1. **Adequação à gramática de casos,** isto é, levar em conta como os vocabulários se adaptam a ONA.
2. **Difusão:** se valoriza o número de vocabulários de LOV que reutilizam os vocabulários anteriores, embora nem todos sejam aqui recolhidos. O LOV constitui uma boa medida, mas não é determinante. Nesses casos, se é possível, é considerada a informação proporcionada pela web do vocabulário.
3. **Tamanho:** número de elementos que compõem o vocabulário. Parece lógico assumir que, a maior quantidade de elementos, menor necessidade de definir classes novas, relações, atributos, etc.
4. **Nível de abstração:** o modelo tenta representar uma realidade complexa. Por esta razão, o nível de abstração do vocabulário a reutilizar não deve agregar mais complexidade. Leva-se em conta o número de elementos que compõem o vocabulário e a complexidade das suas relações.
5. **Outros aspectos,** como a disponibilidade para obter a ferramenta. Foram rejeitadas aquelas ontologias não disponíveis para a sua reutilização (difusão = 0), porque o seu uso está restringido a um âmbito interno, porque não continuaram a ser desenvolvidas, ou bem porque o domínio principal do vocabulário ou o seu âmbito geográfico não se adapta ao analisado.

Como resultado desta análise, foram selecionados onze vocabulários que, por causa das limitações desta memória não são descritos com detalhe. Ficam disponíveis em espanhol no Anexo A. DESCRIPCIÓN DE VOCABULARIOS MAPEADOS. Estas ferramentas são: Schema.org, DUL, Event, FOAF, GEO, GEOP, BBCORE, SNAP, POWER e PROLES.

O mapeamento é iniciado com os tipos de relações entre as ontologias e as categorias básicas dos padrões de correspondência entre estas de Shvaiko e Euzenat (2005) e Scharffe et al. (2014). Quando se integram várias ontologias, podem aparecer distintas incompatibilidades (Bruijn et al., 2006). Habitualmente deve-se a diferenças em relação com o significado ou à codificação dos conceitos em diferentes ontologias; em relação com à linguagem ou com o formalismo empregado para a modelagem do vocabulário (Scharffe et al., 2014). Esta pesquisa foca no primeiro tipo, com as seguintes correspondências:

1. **Correspondência tipo 1:** se dá entre classes que representam o mesmo conceito.
2. **Correspondência tipo 2:** as classes do vocabulário analisado podem ser importadas ao modelo como subclasses ou superclasses, dependendo de cada caso.
3. **Correspondência tipo 3:** a união das classes do vocabulário equivale parcialmente à classe de ONA e podem ser importadas como subclasses desta, cobrindo totalmente sua extensão.
4. **Correspondência tipo 4:** a interseção das classes do vocabulário equivale parcialmente à classe de ONA, que teria outras subclasses.

Quanto às relações, de forma similar às classes, também foram analisadas para o seu mapeamento, identificando-se os mesmos tipos de correspondência entre elementos.

Se o mapeamento descrito nesta memória foi realizado ao nível de ONA, a propriedade *rdfs:isDefinedBy* é utilizada tanto para ONA quanto para ONAPE para a definição dos seus elementos. Portanto, a partir deste processo são dados os primeiros passos do mapeamento dos vocabulários. Também é enfatizada a necessidade de uma análise em detalhe das correspondências identificadas.

3.5. AVALIAÇÃO DE ONA E ONAPE

A avaliação de uma ontologia compreende dois processos: a verificação e a validação.

Neste caso, a **verificação** é realizada em Protégé, por meio do *reasoner* Hermit, aplicado tanto para ONA quanto para ONAPE. Em nenhum dos casos se identificaram erros.

Em relação à **validação**, a análise da adequação de ONAPE (e de ONA) para o seu propósito é efetuada em duas etapas: o povoamento (também denominado de instanciação) de ONAPE e a marcação semântica de notícias com ONAPE. Este processo de avaliação

constitui o início da aprendizagem que requer uma máquina para a obtenção de novas instâncias de ONAPE assim como a descrição de conjuntos de notícias. Neste sentido, é prevista uma pesquisa para o futuro, relacionada com a aprendizagem automática para o desenvolvimento e o uso de ONA.

Seguidamente, é descrito como foi o desenvolvimento do povoamento de ONAPE.

3.5.1. INSTANCIACÃO DE ONAPE

O povoamento da ontologia foi realizado a partir das palavras-chave extraídas do código fonte das notícias da sub-amostra₂. As palavras-chave formam parte do conjunto de metadados que descrevem e facilitam o armazenamento e recuperação de peças de informação (Troncy, 2008).

Como existem múltiplos esquemas de metadados para meios de comunicação, previamente a validação, foram identificados quais esquemas são utilizados pelos jornais, assim como quais são os padrões de descrição adotados por eles. Para esta análise, o levantamento de dados (três notícias por meio) foi realizada em dezembro de 2014 e fevereiro e abril de 2015. Como consequência foram identificados distintos graus de normalização, tanto no uso dos esquemas quanto na forma de inclusão de palavras-chave.

Com base nesta análise foi desenvolvido o aplicativo MetadadosHTML. Este identificou 3,846 termos únicos. A partir destes termos foram selecionados 5 temas:

- **Tema 1:** disputa territorial das Ilhas Senkakus;
- **Tema 2:** homenagem a Nelson Mandela depois de sua morte;
- **Tema 3:** restrições à entrada de imigrantes no Reino Unido e nos Estados Unidos de América;
- **Tema 4:** caso de corrupção na Turquia;
- **Tema 5:** mercado energético.

De cada tema foram selecionadas várias notícias, até obter um total de 23 notícias. Depois, a ONAPE foi instanciada com as palavras-chave destas informações. Seguidamente, com base em Baños-Moreno (2013), foram identificados os tipos de relações entre palavras-chave e elementos de ONAPE:

- **Relação 1** (*eUm*): as palavras-chave são instâncias que encontram sua classe em ONAPE;
- **Relação 2** (equivalência exata): as palavras-chave e as classes da ontologia têm o mesmo significado;

- **Relação 3** (hierarquia de classe): existe uma relação de hierarquia de classe entre as palavras-chave e as classes de ONAPE, bem como de classe-subclasse (tipo 3.1) ou de subclasse-classe (tipo 3.2);
- **Relação 4** (hierarquia de parte): existe uma relação de hierarquia tipo parte-todo entre as palavras-chave e as classes de ONAPE, bem de todo-parte (tipo 4.1) ou de parte-todo (tipo 4.2);
- **Relação 5** (associativa): existe algum tipo de relação semântica entre as palavras-chave e as classes de ONAPE, que são distintas às prévias;
- **Relação 6** (inexistente): não se identificou nenhum tipo de relação entre a palavra chave e qualquer classe da ontologia.

A Tabela 11 traz o total de relações identificadas entre palavras-chave da sub-amostra₂ e elementos de ONAPE, classificadas por temas.

	Tema 1	Tema 2	Tema 3	Tema 4	Tema 5	TOTAL
Relação 1	22	24	11	10	19	86
Relação 2	5	3	7	0	7	22
Relação 3	0	0	0	0	1	1
Relação 4	0	1	5	0	2	8
Relação 5	0	0	0	0	2	2
Relação 6	0	1	0	0	1	2
TOTAL	27	29	23	10	32	121

Tabela 11: Total de relações entre a sub-amostra₂ e elementos de ONAPE por tema

No total foram obtidas 121 palavras-chave diferentes para os cinco temas analisados. Destas, mais de **71%** são instâncias para as quais ONAPE tem classes específicas. Um percentual de **18,2%** são ou equivalem a classes levantadas pela ontologia. Isto significa que **89%** das palavras-chave que aparecem no código fonte das notícias podem ser representadas por meio de elementos da ontologia. Por outro lado, mais de **9%** de palavras-chave mantêm algum tipo de relação com elementos de ONAPE, sejam do tipo classe-subclasse, todo-parte ou associativas. Ou seja, para mais de **98%** das palavras-chave foi encontrado algum tipo de relação. As exceções são os termos *Qualidade do ar* e *Legado de Mandela*, que representam aproximadamente mais de **1,6%** sobre o total de termos.

Além dos dados anteriores, para cada tema foram calculadas: a exaustividade, a precisão e a medida-F, tanto exatas quanto aproximadas. Estas são mostradas na Tabela 12.

	Tema 1	Tema 2	Tema 3	Tema 4	Tema 5	MÉDIA
Exaustividade (exata)	1	0,93	0,78	1	0,81	0,90
Precisão (exata)	1	0,96	0,78	1	0,84	0,91
Medida-F (exata)	0,50	0,47	0,39	0,50	0,41	0,45
Exaustividade (aproximada)	1	0,93	1	1	0,91	0,96
Precisão (aproximada)	1	0,96	1	1	0,94	0,98
Medida-F (aproximada)	0,5	0,47	0,50	0,50	0,46	0,49

Tabela 12: Exaustividade, Precisão e medida-F por temas analisados

Estes dados mostram índices elevados de representatividade das notícias por ONAPE, pois em todos os casos foi obtido um resultado superior a **78%** de exaustividade e precisão (**100%** para os temas 1 e 4), surgindo deste modo os valores da medida harmônica (medida-F). Os resultados são melhores com a inclusão nos cálculos as relações 3.2 (classe-subclasse) e 4.2 (todo-parte), particularmente para o tema 3. Assim, a exaustividade e precisão oscilam de 0,78 a 1. Consequentemente a medida-F tem um melhor resultado, de 0,45 a 0,49. Para conhecer como foram calculados estes índices pode-se visualizar as Figuras 24, 25, 26.

3.5.2. MARCAÇÃO SEMÂNTICA DE NOTÍCIAS COM ONAPE

Nesta segunda avaliação se marcaram semanticamente as notícias do tema 1 da submostra₂: **Disputa territorial das Ilhas Senkaku**. Desta forma, foi identificada cada entidade básica (ou classe de ONAPE) por as cores indicadas na Tabela 32 em 4.5.4. VALIDACIÓN POR ANOTACIÓN SEMÁNTICA. As notícias, marcadas empregando estas cores, são mostradas no Anexo C. MARCADO SEMÁNTICO DE NOTÍCIAS DEL TEMA 1. Os resultados deste processo são indicados na Tabela 13.

ONA	CLASSES DE ONAPE	F. INSTÂNCIAS ÚNICAS	F. TOTAL INSTÂNCIAS
Ações	44	103	200
Agentes	15	75	199
Coisas	41	72	112
Espaços físicos	7	31	99
Espaços temporais	7	25	35
TOTAL CLASSES	114	306	645
Propriedades	12	64	80
TOTAL PROPRIEDADES	126	370	725

Tabela 13: Total de elementos do tema 1 da sub-amostra₂ identificados em ONAPE

RESUMO EM PORTUGUÊS

A primeira coluna apresenta as principais classes de ONA e ONAPE, enquanto na segunda faz-se referência à frequência (F.) de classes únicas (diferentes) identificadas nos textos do tema 1. a segunda coluna indica a frequência de instâncias únicas (diferentes). Já última refere a frequência total de instâncias (com repetições).

Identificaram-se primeiramente ações (44), seguidas de coisas (41) e agentes (15). Ações e coisas representaram 67,4% do total de elementos. Da mesma forma, no caso das instâncias únicas se encontraram mais ações (103), seguidas por agentes (75) e coisas (72). A soma destes três tipos representam mais de 67% do conjunto de instâncias únicas. As ações supõem, de fato, quase 27% do total dos elementos. Por último, quanto à frequência total de instâncias, praticamente se identificou a mesma quantidade de ações do que agentes, com uma taxa de repetição de 1,94 e 2,65, respetivamente. Este dado ressalta especialmente para os espaços físicos que, com 31 instâncias únicas, multiplica por 3 a taxa média de repetição (3,19). Isto é devido às contínuas remissões aos territórios em conflito: ilhas Senkakus ou Diaoyous e mar da China Meridional, entre outros. Em quanto às propriedades, tanto de tipo dado quanto de tipo objeto, foram identificados 12 tipos, com 64 instâncias únicas e 80 em total, o que indica uma baixa frequência.

A Tabela 14 mostra o total de relações estabelecidas entre classes e instâncias únicas e elementos de ONAPE por notícias. Foi considerado um total de 323 elementos diferentes para as cinco notícias analisadas. Mais de 94% são instâncias para as quais ONAPE contempla uma classe mais ou menos específica, 3% são ou equivalem a classes da ontologia. Isto significa que 97% das palavras-chave podem ser representadas por ONAPE. Oito elementos fazem a diferença (2% do total), para os quais não se encontraram classes suficientemente específicas nem classes equivalentes. Quanto às propriedades, também se identificou uma nova propriedade para ONAPE, *alias (pen name)*. Esta propriedade, subclasse de *nome*, pode-se atribuir a quase qualquer elemento e particularmente a pessoas.

	Notícia 1	Notícia 2	Notícia 3	Notícia 4	Notícia 5	TOTAL
Relação 1	3	1	1	1	4	10
Relação 2	61	28	44	56	117	306
Relação 3	0	0	0	0	0	0
Relação 4	0	0	0	0	0	0
Relação 5	0	0	0	0	0	0
Relação 6	0	4	1	2	0	7
TOTAL	64	33	46	59	121	323

Tabela 14: Relações entre términos de notícias e elementos de ONAPE por notícia

Por último, para cada notícia foi calculada exaustividade, precisão e medida-F exatas (ver 4.5.4. VALIDACIÓN POR ANOTACIÓN SEMÁNTICA). Esses dados mostram taxas elevadas de exaustividade e precisão, assim como de harmonia entre elas (medida-F). Os dados são melhores que na primeira validação, provavelmente porque existe maior homogeneidade de termos entre as notícias analisadas. Em todos os casos foi obtido um resultado superior ao **88%** de exaustividade e precisão, sendo do **100%** para as notícias 1 e 5.

	Notícia 1	Notícia 2	Notícia 3	Notícia 4	Notícia 5	MÉDIA
Exaustividade (exata)	1	0,88	0,98	0,97	1	0,96
Precisão (exata)	1	1	1	1	1	1
Medida-F (exata)	0,5	0,47	0,49	0,49	0,5	0,49

Tabela 15: Exaustividade, Precisão e medida-F por temas analisados

Neste caso não procede o cálculo dos índices aproximados, pois não se identificaram relações do tipo 3 (hierarquia de classe) nem se levaram em conta as relações de tipo 4 (hierarquia de parte).

3.6. ESPECIFICAÇÃO DE ONA E ONAPE

A última etapa de construção de uma ontologia é a criação da sua documentação, em que esta é descrita. Nesta pesquisa foi utilizado *Live OWL Documentation Environment* (LODE) para gerar automaticamente a documentação tanto de ONA quanto de ONAPE. No Anexo Error: no se encontró el origen de la referencia consta a documentação da ontologia ONA. Devido às limitações desta resumo não é possível mostrar a especificação de ONAPE.

4. CONCLUSÕES E PROPOSTAS

Esta seção é uma tradução completa para o português do capítulo 5. CONCLUSIONES, PROPUESTAS Y CONTRIBUCIONES desta tese. Nele destacam-se as principais conclusões alcançadas, tanto na pesquisa desenvolvida quanto nos produtos obtidos, ONA e ONAPE. Também se descrevem alguns trabalhos futuros, que constituem possíveis linhas de pesquisa nas quais os modelos são o ponto de partida. Além disso, existe uma última seção de contribuições, que se encontram referenciadas na alínea 5.3. PUBLICACIONES CIENTÍFICAS e não precisam de tradução para o português.

4.1. CONCLUSÕES

Nenhuma organização pode pretender que um modelo se ajuste para qualquer tipo de coisa nem para todos os casos, sequer o Google conseguiu esse fato (Stuart, 2016, p. 75). Contudo, tem-se procurado desenvolver vocabulários gerais e reutilizáveis em diferentes

contextos, sejam peças de informação ou outro tipo de documentos. O marco teórico, baseado principalmente nos aportes teóricos de Fillmore (1967) e García Gutiérrez (2014), revelam-se úteis para o propósito da modelagem da ontologia geral ONA, a partir da qual criou-se a ontologia de domínio, ONAPE. Os dois vocabulários foram avaliados em duas etapas. Primeiramente com as palavras-chave empregadas pelos próprios meios de comunicação, que consideraram-nas como termos representativos dos seus conteúdos; depois, por meio da análise dos textos completos de uma amostra de notícias menor. Em ambos os casos, foram obtidos altos níveis de exaustividade e precisão, com uma média harmônica (a medida-F). Isto constata tanto a adequação dos termos de ONAPE para a representação das peças de informação (precisão) quanto do seu nível de especialização (exaustividade).

Não obstante, é importante reconhecer que uma ontologia é necessariamente uma representação parcial e imperfeita de algo e não uma reflexão completa do mundo real. Escolher [e modelar] uma ontologia implica a decisão de focar em aspetos concretos desse mundo (Stuart, 2016, p. 83). Além disso, como todo vocabulário, e especificamente como toda ontologia, a ONAPE aporta certa visão do mundo para o domínio analisado, como já indicaram Uschold e Grüninger (1996).

Levando em conta estas questões e tendo em vista os resultados das avaliações de ONAPE, a hipótese submetida no início deste documento é aceita. Em outras palavras, é possível a modelagem de uma ontologia de domínio político-econômico por meio da análise de uma amostra de notícias de meios de comunicação impressa baseada na indexação, tanto humana quanto automática.

Em relação aos aspetos concretos da pesquisa, seguem algumas conclusões:

1. Os modelos desenvolvidos nesta memória, ONA e ONAPE são produto da simbiose de duas áreas: Ciência da Informação e Informática. De um lado, foram aproveitadas algumas ferramentas criadas por profissionais da Computação (eTermos, Grampal, MetadadosHTML); por outro, se aplicaram técnicas de análise e indexação, próprias da Ciência da informação, sobre um conjunto de notícias. Os processos intelectuais precisaram de um investimento de tempo maior, vinculado estreitamente com a complexidade do domínio político-econômico, voltados para o contexto internacional dos meios de comunicação. Isto acontece porque cada país possui uma realidade diferente que, depois, é descrita por um jornal. Finalmente essa realidade é percebida por pessoas. Por esta razão é difícil interpretar o que se diz e o que se quer dizer, mesmo para outra pessoa. A experiência prévia e os conhecimentos adquiridos nesta pesquisa têm sido essenciais para esta análise.

2. Em relação com a fundamentação metodológica, esta investigação contribui para a visualização das relações entre distintos vocabulários (taxonomias, tesouros, ontologias), desenvolvidos para a representação do conteúdo dos documentos (representação semântica). A ontologia, seguindo Codina e Pedraza-Jiménez (2017), entre outros, é apresentada como uma evolução das ferramentas que vêm aparecendo ao longo da história, de acordo com as necessidades de recuperação de informação. Esse vínculo é importante, pois permite aproximar duas posturas que, à vista da bibliografia utilizada, continuam a se manter separadas: vocabulários desenvolvidos na área de Ciência da Informação, vocabulários desenvolvidos na área da Computação.
3. No que se refere as metodologias de modelagem de ontologias e ao alvo 1 desta tese, como indicam Mendonça e Almeida (2014) e Silva (2008), existem múltiplas referências. Mesmo assim, nenhuma delas conseguiu se converter em um padrão, motivo pelo qual se optou em seguir os princípios básicos de modelagem identificadas por estes autores. Esta perspectiva se adapta, em termos gerais, à criação de qualquer outro vocabulário, especialmente tesouros, que são, provavelmente, os mais parecidos com as ontologias. Embora não se tenha mantido a nomenclatura utilizada pelos autores, são mais ou menos seguidos ao longo da memória. Sobre esta questão, no desenho da metodologia, convém destacar a importância do uso da linguagem e de um grau de descrição que facilite seu entendimento. Os vocabulários nem sempre serão desenvolvidos nem utilizados por documentalistas ou profissionais de TI.
4. Com a indexação humana e semiautomática, empregadas para a atribuição, no primeiro caso, e a extração, no segundo, de termos de indexação com vistas à dotação de classes e relações para ONAPE, conseguiu-se um grande número de elementos. Esta pesquisa, além disso, é consistente com alguns aspetos já apontados por García Gutiérrez (2014) no que tem a ver com a capacidade representativa dos títulos das notícias: muitas vezes “há uma implacável retórica dentro de uma tipologia textual muito extensa”. Entre elas destacam-se o uso de orações impessoais, orações transitivas interrompidas ou orações com omissão verbal. Por esta razão, incrementar o tamanho dos textos de análise pode proporcionar uma maior quantidade de potenciais elementos. O ruído que esta nova aplicação ofereça no processo deve ser analisado para avaliar os seus custos em relação à qualidade do produto. O Processamento da Linguagem Natural (PLN), portanto, tem sido utilizado tanto para extrair termos para a modelagem de uma ontologia quanto para aplicar os termos destas ferramentas para indexação de documentos. A colaboração com profissionais da TI é precisa para isto (ver 5.2. TRABAJOS FUTUROS).

5. A respeito do alvo 2 desta tese, existem precedentes de indexação automática em documentos jornalísticos, por exemplo, García et al. (2006) ou Maidel et al. (2010). Todavia, a maioria usa taxonomias gerais da IPTC e não analisam um domínio concreto nem focam nas ações descritas nas notícias. Quanto ao domínio político-econômico, foram encontrados trabalhos sobre ontologias especializadas, como Ortiz (2011) ou Daquino et al. (2014), mas focam em aspetos distintos às ações ou abordam procedimentos diferentes. O Anexo A.1. HERRAMIENTAS REUTILIZADAS POR LOS VOCABULARIOS traz descrição mais detalhada em relação às características de algumas ontologias analisadas, gerais e especializadas em média escrita e política-economia, com vistas a uma reutilização posterior.
6. Em relação ao esquemas de metadados e ao alvo 3, foram analisados tanto esquemas gerais quanto os aplicados para os meios de comunicação escrita e os seus documentos. Isto permitiu uma aproximação em relação aos metadados de conteúdo (semânticos). Em Baños-Moreno et al. (2015) se destaca, para o caso dos jornais, a convivência de diversos padrões de metadados. Estes dois produtos, esquemas de metadados e ontologias focam em vários problemas: a fugacidade da informação nas redes, a multiplicação da informação e dos vocabulários desenhados para seu melhor controle. Neste sentido, em Baños-Moreno et al. (2017) se colocam em evidência que não somente existe um problema de quantidade mas de qualidade da descrição da informação, com práticas de representação de conteúdos não sistemáticas nem compartilhadas, o que dificulta a reutilização de informação. Por esta razão, tem se previsto completar o mapeamento de ONA e ONAPE com outros vocabulários analisados (ver 5.2. TRABAJOS FUTUROS). Assim, no futuro, elementos das ontologias serão substituídos por outros de vocabulários que foram definidos como fonte de definição de classes e propriedades. Além disso, as duas ferramentas foram disponibilizadas no *site* da Universidad de Murcia com a sua documentação correspondente. Desta forma, pretende-se facilitar o acesso e a reutilização delas, respeitando os princípios indicados por Mendonça e Almeida (2014).
7. A respeito de MetadadosHTML e levando em conta o alvo 3 deste tese, destaca-se a sujeição as mudanças nos códigos fonte dos meios de comunicação. Se um jornal redefine alguma das suas etiquetas meta ou começa a utilizar um esquema novo, o aplicativo não pode aproveitá-las, o que obriga a uma contínua revisão. Além disso, é importante sublinhar a dificuldade para obter todos os atributos que aparecem nos códigos fonte, pois não sempre estão disponíveis. Diante disto, percebe-se a necessidade de definir uma política de indexação alinhada com os padrões de metadados. Ficou evidenciado que a simples adoção de metadados não é suficiente para a difusão e acessibilidade dos documentos.

8. Sobre a atualização do vocabulário de representação ONAPE, ainda mais no domínio tão amplo como é o político-econômico, deve ser revisado. Isto implica a inclusão dos novos aspetos não abrangidos até agora e, talvez a supressão de outros. Para efeito, é preciso analisar novos *corpora* de notícias (ou outros tipos de documentos) e, como destaca Hlomani (2014), adaptá-lo em cada período de análise. Somente desta forma é possível assegurar a validade de ONAPE no tempo. Em outras palavras, que responda as necessidades de descrição das novas realidades. Tal como foi submetida a construção das duas ontologias, mesmo como sua avaliação, o alvo 4 da tese, esta pesquisa constitui o ponto de partida para um desenvolvimento maior. Assim, é que se propõe como pesquisa para o futuro a atualização da ontologia por meio da aprendizagem baseada na experiência, como é referido em 5.2. TRABAJOS FUTUROS.
9. Finalmente, com respeito à evolução da web sintática (Romero Llop, 2007) para a web semântica, é um caminho dinâmico no que surgem e devem continuar a surgir técnicas novas e vocabulários orientadas ao tratamento da informação. Essas ferramentas facilitarão a marcação semântica para um volume de conteúdos disponíveis no Internet cada vez maior. Para isto, os especialistas em informática criam técnicas e algoritmos. Não obstante, ainda é preciso a participação ativa dos(as) profissionais da Ciência da Informação, sejam documentalistas, arquivistas, bibliotecários(as), etc., que têm a experiência e a formação para melhorar os resultados dos produtos da Computação. Tennis e Calzada-Prado (2007) já destacavam o campo de possibilidades que se abria para as profissões da Ciência da Informação a este respeito. São elas quem se defrontam diariamente com os problemas de acesso à informação, seja por causa de uma descrição insuficiente ou por carências da alfabetização digital. Desde o ponto de vista formativo, é necessário incrementar a carga letiva dos graus de Ciência da Informação em tecnologias da Web Semântica. Entender as estruturas das linguagens de marcação, como HTML, do XML, da linguagem RDF ou da OWL, também pode contribuir para melhor análise dos vocabulários em linguagem formal e suas possibilidades. Além disto, aprender utilizar as diversas ferramentas de processamento automático da informação pode reduzir as tarefas de caráter mecânico e dispor de mais tempo para atividades intelectuais.

4.2. PROPOSTAS PARA O FUTURO

Nesta seção são relacionadas propostas de trabalhos futuros. Estas se referem à ontologia ONA e sobre tudo, a ONAPE.

MAPEAMENTO COM VOCABULÁRIOS REUTILIZÁVEIS

Como já tem sido indicado anteriormente, esta proposta consiste em completar o mapeamento, tanto de ONA quanto de ONAPE, com elementos de outros vocabulários, descritos em 4.4.3. MAPEO DE VOCABULARIOS. O processo é iniciado nesta pesquisa por meio da inclusão de propriedades de tipo anotação (*isDefinedBy*) para especificar as fontes de definição de determinadas classes e propriedades das duas ontologias. Este trabalho futuro focaria na substituição desses componentes por elementos importados do vocabulário de origem.

ESPECIALIZAÇÃO EM SUBDOMÍNIOS

O domínio político-econômico é amplo, portanto é possível submeter o desenvolvimento de subdomínios de interesse, tendo em conta as necessidades informativas que possam surgir no futuro. Por exemplo, as atividades codificadas em *Codificación Nacional de Actividades Económicas* (CNAE)⁵⁴ podem ser úteis para detalhar aspectos concretos das ações *AçãoAtividadeEconómica*, criando novos elementos ou completando definições, descrições e relações desta classe.

Outro aspeto mais interessante tem a ver com a corrupção, política ou econômica. Não são poucas as notícias que diariamente informam de mais casos de corrupção. Um melhor controle da informação neste tema pode facilitar a tomada de decisões pela cidadania, por exemplo, nos processos eleitorais. Também pode ser útil para outro tipo de usuários. Assim, jornalistas e documentalistas de meios de comunicação poderiam recuperar informação mais precisa sobre um caso concreto, por meio da marcação de documentos com termos da ONAPE mais especializados. Tal é o interesse nesta questão que já existem precedentes de análise de documentos deste tema com ontologias, como Corrupteca (ver 4.4.1. PRESELECCIÓN DE VOCABULARIOS).

ATUALIZAÇÃO DE ONAPE POR MEIO DA ANÁLISE DE NOVOS CORPORA

Para o futuro seria interessante a inclusão de novos elementos, que permitam modular as ações, estabelecendo graus de execução destas ou bem detalhando o que acontece. Por exemplo, pode-se submeter uma notícia, relativamente frequente nos jornais: sobem as vendas do produto x. A ação propriamente dita não é uma venda senão um conjunto de vendas. Para contextualizar melhor essa peça de informação se definiria a propriedade,

⁵⁴ CNAE: <http://www.cnae.eu/> (último acesso: 18/05/2017).

temModulador, que relacionaria uma *ação* e uma subclasse de *ModuladorDeAção*. Desta forma uma *AçãoVenta* teria *Sobida* como modulador de dita ação.

Algumas destas subclasses podem ser definidas das seguintes formas:

- **Baixa:** indica a queda ou diminuição de algo. Por exemplo, a queda do número de usuários de um produto. Termos que também refletem a ideia são *Redução*, *Diminuição*, *Declive*, *Decrescimento*, *Queda*, *Colapso* e *Descida*.
- **Estagnação:** permite exprimir a manutenção das quantidades ou cifras de execução de alguma ação.
- **Fixação:** classe para exprimir o estabelecimento de algum elemento em qualquer tipo de *Ação*, como o *Preço* de um *Produto* ou o *Salário* de trabalhadores.
- **Tentativa:** tentativa e fracasso de conseguir algo através da execução de uma *Ação*.
- **Manutenção:** classe que exprime a manutenção de alguma condição ou situação.
- **Subida:** classe pela qual se exprime o aumento de algo, em alguns tipos de ações. Pode-se utilizar, por exemplo, para exprimir o crescimento das vendas de um produto. Termos que refletem a ideia são *Aumento*, *Crescimento*, *Expansão* e *Incremento*. Também *Escala* e *Escalada* se empregar para refletir o incremento ou subida de algo, como *Escalada de Tensão*.

Contudo, a manutenção e sustentabilidade das ontologias são complicadas. A sua criação já consome muito tempo e este persiste se quiser disponibilizar e mantê-la atualizada (Stuart, 2016, p. 125). Por isto um trabalho para o futuro estabelece um projeto de instanciação (e atualização) de ONAPE por meio da aprendizagem baseada na experiência. Esta técnica prevê a automatização da instanciação da ferramenta levando em conta uma etapa prévia de marcação semântica humana. Desta forma, o sistema aprende não somente os termos que são instâncias, senão que estruturas se repetem. Por exemplo: determinante + substantivo + adjetivo + verbo.

Quanto aos *corpora* de análise, é possível recorrer a várias opções na marcação das peças de informação: 1) ampliar o tamanho do texto a analisar, pois segundo Guerrillot (2006), o assunto principal da notícia se encontra entre as 100 primeiras palavras, em um ou dois parágrafos; 2) construir macroproposições, no sentido que exprime García Gutiérrez (2014): “síntese textual elaborada a partir da formulação das 12W para cada notícia da amostra, que consiste em *reescrever* (e, no seu caso, substituir) e enriquecer o título em um mesmo enunciado canônico do tipo: sujeito-ação-predicado com os dados que o próprio texto original explicitamente proporcione”.

Outra possibilidade é a utilização dos conteúdos produzidos no contexto dos meios de comunicação, mas não por jornalistas, senão pelos leitores, tais como *posts*, comentários, etc. Estes, convertidos em *prosumidores* (Toffler, 1980) participam ativamente na geração de conteúdos graças a ferramentas como blogs, comunidades ou comentários, dentro e fora das webs dos jornais. Assim, *The New York Times*, *The Guardian* e a BBC, entre outros, tem começado a publicar alguns dos seus conteúdos diretamente nas redes sociais (Jiménez Cano e Abad Liñán, 2015). Seu controle e recuperação serão fundamentais para responder adequadamente às necessidades informativas da cidadania.

Por último, também se considera recorrer a entrevistas com especialistas, tanto dos domínios político-econômico quanto do jornalismo e Ciência da Informação, com vistas a melhorar da ferramenta: que problemas encontram os pesquisadores e profissionais para a recuperação de documentos? Como gostariam de encontrar um assunto hipotético em determinado sistema de recuperação de informação?

Uma questão adicional tem a ver com a inclusão de sinônimos, enriquecendo assim as possibilidades de marcação semântica de documentos. Para isto, se analisariam fontes ricas em termos, como Eurovoc, e ontologias linguísticas, como Wordnet.

MELHORIA DA AVALIAÇÃO DAS ONTOLOGIAS

Outro aspeto no que trabalhar no futuro é uma maior aprofundização da análise das ontologias com as medidas de avaliação empregadas: exaustividade, precisão e medida-F. Para isto, se utilizaria uma amostra maior de peças de informação de forma que facilite a análise de dados agregados. Igualmente é possível o uso de ferramentas especializadas na avaliação de ontologias.

MELHORIA DE METADADOSHTML

Os resultados iniciais mostram a necessidade de estabelecer uma ordem de preferência na extração de valores de dados a partir de uns esquemas ou outros. Para tal, MetadadosHTML será modificado levando em conta a frequência no uso dos esquemas e o grau de padronização das práticas de cada meio. A Tabela 16 indica a ordem de extração de palavras-chave empregada.

Ordem	Linha de código	Meios	Nota
1	<meta name="keywords" content="">	101; 121; 123; 202; 311; 401; 601; 602; 603; 701; 711; 721	Para 201 se utiliza 4, 5, 6 e 7 Campo repetível em 701
2	<meta name="news_keywords" content="">	122; 731; 801	Para 201 se utiliza 4, 5, 6 e 7
3	<meta property="article:tag" content="" />	301; 901	Campo repetível
4	<meta name="des" content="" />	201	palavras-chave
5	<meta name="org" content="" />	201	Organizações
6	<meta name="per" content="" />	201	Pessoas
7	<meta name="geo" content="" />	201	Lugares

Tabela 16: Informação sobre o metadado "palavras-chave" para MetadadosHTML

O propósito é assegurar a correta extração de valores de dados com o menor número de processos. A informação extraída será utilizada para modificar ONAPE com mais classes, subclasses e relações.

*

**

BIBLIOGRAFÍA

- Abadal, E., & Codina, L. (2009). *Bases de datos documentales. Características, funciones y método*. Madrid: Síntesis.
- Abadal, E., Guallar, J., & Codina, L. (2014). Sistemi di documentazione della stampa periodica: quali sono e come valutarli? *AIB studi*, 54(1), 75-86. <https://doi.org/10.2426/aibstudi-9486>
- Abbas, J. (2010). *Structures for Organizing Knowledge: Exploring Taxonomies, Ontologies, and Other Schemas*. New York: Neal-Schuman Publishers.
- Abbud Grácio, J. C., & Fadel, B. (2010). Estratégias de preservação digital. En *Gestão, mediação e uso da informação* (pp. 58-83). São Paulo: Editora UNESP; Cultura Acadêmica. Recuperado a partir de <http://books.scielo.org/id/j4gkh/pdf/valentim-9788579831171-04.pdf>
- Acurero, A., Bracho, D., & Rincón, C. (2008). Análisis Preliminar sobre la Aplicación de Ontologías al Descubrimiento de Servicios Web Semánticos. Presentado en Séptima Conferencia Iberoamericana en Cibernética e Informática: CISCI 2008, Maracaibo, Venezuela.
- AENOR. (1990, diciembre 13). Norma UNE 50106:1990. Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües. Recuperado a partir de http://indizacionytesauros.wikispaces.com/file/view/Norma+50106_90.pdf
- AENOR. (2014, noviembre 26). UNE-ISO 25964-1:2014 Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 1: Tesauros para la recuperación de la información. AENOR. Recuperado a partir de <http://www.aenor.es/aenor/normas/normas/fichanorma.asp?tipo=N&codigo=N0053960>
- Agarwal, S., Singhal, A., & Bedi, P. (2012). Classification of RSS Feed News Items Using Ontology. En A. Abraham, A. Zomaya, S. Ventura, R. Yager, V. Snasel, A. K. Muda,

- & P. Samuel (Eds.), *International Conference on Intelligent Systems Design and Applications*, ISDA (pp. 491-496). New York: Ieee.
<https://doi.org/10.1109/ISDA.2012.6416587>
- Aguado-de-Cea, G., Montiel-Ponsoda, E., Poveda-Villalón, M., & Giraldo-Pasmin, O. X. (2015). Lexicalizing Ontologies: The Issues Behind The Labels. *Procedia - Social and Behavioral Sciences*, 212, 151-158.
<https://doi.org/10.1016/j.sbspro.2015.11.313>
- Aibar, E., Dunajcsik, P., Lerga, M., Lladós, J., Meseguer, A., & Minguillón, J. (2016). Guía de recomendaciones y buenas prácticas para editar el contenido científico de Wikipedia. FECYT; UOC. Recuperado a partir de <http://hdl.handle.net/10609/51462>
- Aibar, E., Minguillón, J., Lerga, M., Lladós, J., Meseguer, A., & Dunajcsik, P. (2016). *Análisis del contenido científico de la Wikipedia española*. FECYT; UOC. Recuperado a partir de http://openaccess.uoc.edu/webapps/o2/bitstream/10609/51462/1/Gu%C3%ADa_Wikipedia_Ciencia.pdf
- Akarcesme, S. (2016, marzo 8). Recep Tayyip Erdogan's Despotic Zeal. *The New York Times*. Recuperado a partir de <https://www.nytimes.com/2016/03/09/opinion/recep-tayyip-erdogans-despotic-zeal.html>
- Alesso, H. P., & Smith, C. F. (2004). *Developing Semantic Web Services*. Natick, Mass: A K Peters/CRC Press.
- Allemang, D., & Hendler, J. A. (2011). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL* (2a ed.). Amsterdam ; Boston: Elsevier Ltd, Oxford.
- Allen, R. B., Japzon, A., Achananuparp, P., & Lee, K. J. (2007). A framework for text processing and supporting access to collections of digitized historical newspapers. En M. J. Smith & G. Salvendy (Eds.), *Human Interface and the Management of*

- Information: Interacting in Information Environments, Pt 2, Proceedings* (Vol. 4558, pp. 235-244). Berlin: Springer-Verlag Berlin. Recuperado a partir de <https://scholarworks.iupui.edu/bitstream/handle/1805/4552/allen-2007-framework.pdf?sequence=1&isAllowed=y>
- Alvite Díaz, & Martínez González, M. M. (2015). Armonización de ISO 25964 y SKOS en software de gestión de tesauros. Propuesta de análisis. En *Organización del conocimiento: sistemas de información abiertos* (pp. 266-278). Facultad de Comunicación y Documentación, Universidad de Murcia, Murcia (España). Recuperado a partir de http://www.iskoiberico.org/wp-content/uploads/2015/11/43_Ba%C3%B1os.pdf
- Anam, S., Kim, Y. S., Kang, B. H., & Liu, Q. (2016). Adapting a Knowledge-based Schema Matching System for Ontology Mapping. En *Proceedings of the Australasian Computer Science Week Multiconference* (p. 27:1–27:10). New York, NY, USA: ACM. <https://doi.org/10.1145/2843043.2843048>
- Anderson, J. D., & Pérez-Carballo, J. (2001a). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing. *Information Processing & Management*, 37(2), 231-254. [https://doi.org/10.1016/S0306-4573\(00\)00046-7](https://doi.org/10.1016/S0306-4573(00)00046-7)
- Anderson, J. D., & Pérez-Carballo, J. (2001b). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part II: Machine indexing, and the allocation of human versus machine effort. *Information Processing & Management*, 37(2), 255-277. [https://doi.org/10.1016/S0306-4573\(00\)00046-7](https://doi.org/10.1016/S0306-4573(00)00046-7)
- Antoniou, G., & Harmelen, F. van. (2009). Web Ontology Language: OWL. En P. D. S. Staab & P. D. R. Studer (Eds.), *Handbook on Ontologies* (2a ed., pp. 91-110). Berlin, Germany: Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-24750-0_4

- Baños-Moreno, M.-J., Felipe, E. R., Pastor-Sánchez, J.-A., Lima, G., & Martínez-Béjar, R. (2017). Análisis de metadatos de noticias para la extracción de información del código fuente. El software METADATOSHTML. *Information Research*, 22(1, paper 740). Recuperado a partir de <http://www.informationr.net/ir/22-1/paper740.html>
- Baños-Moreno, María-José, Felipe, E. R., Pastor-Sánchez, J. A., Martínez-Béjar, R., & Lima, G. (2015). Metadatos en noticias: un análisis internacional para la representación de contenidos en periódicos. Presentado en II Congreso ISKO España y Portugal / XII Congreso ISKO, Facultad de Comunicación y Documentación, Universidad de Murcia, Murcia (España). Recuperado a partir de http://www.iskoiberico.org/wp-content/uploads/2015/11/43_Ba%C3%B1os.pdf
- Baños-Moreno, M.-J. (2013). Fuentes para la actualización de macro-tesauros: Noticias de divulgación científica. *Cuadernos de Gestión de Información*, 3(1), 13–24.
- Baños-Moreno, M.-J., Pastor-Sánchez, J.-A., & Martínez-Béjar, R. (2017). Interactivity features of online newspapers: from a facsimile model to a multimedia one. *Anales de Documentación*, 20(2), 1-18.
- Benjamins, V. R., & Gómez-Pérez, A. (2000). Knowledge-system technology: ontologies and problem-solving methods. Recuperado a partir de http://www.researchgate.net/profile/V_Richard_Benjamins/publication/265263041_Knowledge-System_Technology_Ontologies_and_Problem-Solving_Methods/links/54acec2a0cf21c47713a773c.pdf
- Bernal Torres, D. M., Castro Romero, A., & González Sanabria, J. S. (2017). Web semántica, más de una década de su aparición. *Puente*, 8(1), 61-69. <https://doi.org/10.18566/v8n1.a07>
- Berners-Lee, T. (2000). *Semantic Web on XML*. Presentado en XML 2000, Washington DC. Recuperado a partir de <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10->

0.html

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284(5), 34-43. <https://doi.org/10.1038/scientificamerican0501-34>
- Beydoun, G., Lopez-Lorca, A. A., García-Sánchez, F., & Martínez-Béjar, R. (2011). How do we measure and improve the quality of a hierarchical ontology? *The Journal of Systems and Software*, 84(12), 2363-2373. <https://doi.org/http://dx.doi.org/10.1016/j.jss.2011.07.010>
- Biblioteca Nacional de Francia. (2014). BnF - Modelo CIDOC CRM. Recuperado 29 de septiembre de 2015, a partir de http://www.bnf.fr/es/profesionales/modelizacion_ontologias/a.modelo_cidoc_crm.html
- Bourdieu, P. (1977). Structures, Habitus, Power: Basis for a Theory of Symbolic Power. En P. Bourdieu, R. Nice (Trad.), *Outline of a Theory of Practice* (pp. 159-197). Cambridge: Cambridge University Press. Recuperado a partir de <http://ebooks.cambridge.org/chapter.jsf?bid=CBO9780511812507&cid=CBO9780511812507A034&tabName=Chapter>
- Brickley, D., & Miller, L. (2014, enero 14). FOAF Vocabulary Specification 0.99. Namespace Document 14 January 2014 - Paddington Edition. Recuperado a partir de <http://xmlns.com/foaf/spec/>
- Bruijn, J. de, Ehrig, M., Feier, C., Martíns-Recuerda, F., Scharffe, F., & Weiten, M. (2006). Ontology Mediation, Merging, and Aligning. En J. Davies, R. Studer, & P. Warren (Eds.), *Semantic Web Technologies: Trends and Research in Ontology-based Systems* (pp. 95-113). England: John Wiley & Sons, Ltd. Recuperado a partir de <http://onlinelibrary.wiley.com/doi/10.1002/047003033X.ch6/summary>

- Carolo, F., & Rocha, L. B. L. da. (2011). Improving Web Content Management with Semantic Technologies. Presentado en 2011 Semantic Technology Conference (SemTech 2011), San Francisco, CA. Recuperado a partir de http://semtech2011.semanticweb.com/uploads/handouts/Carolo_Fernando_3919_3311.pdf
- Castells, P., Perdrix, F., Pulido, E., Rico, M., Fuentes, J. M., Benjamins, R., ... Granollers, T. (2006). Newspaper Archives on the Semantic Web. En R. Navarro-Prieto & J. L. Vidal (Eds.), *HCI related papers of Interacción 2004* (pp. 267-276). Springer Netherlands. Recuperado a partir de http://link.springer.com/chapter/10.1007/1-4020-4205-1_22
- Cavalcanti, C. R. (1978). *Indexação & tesouro: metodologia & técnicas* (preliminar). Brasília - DF, Brazil: Associação dos bibliotecários do distrito federal (ABDF).
- Chowdhury, G. G. (2010). *Introduction to modern information retrieval* (3rd ed.). New York, NY, USA: Neal-Schuman Publishers.
- Cleveland, D. B., & Cleveland, A. D. (2013). *Introduction to indexing and abstracting* (4th ed.). Santa Barbara, California: Unlimited libraries.
- Codina, L. (2010). El periodismo digital: lo mejor que tenemos en la web (epílogo). En *Prensa digital y bibliotecas* (p. 170). Trea. Recuperado a partir de http://www.lluiscodina.com/periodismoDigital_2010.pdf
- Codina, L. (2017, enero). Componentes estructurales y semánticos en el diseño de la navegación Web: taxonomías, SEO y software implicado en el caso de una instalación con WordPress. Recuperado a partir de <http://repositori.upf.edu/handle/10230/27879>
- Codina, L., & Pedraza-Jiménez, R. (2011). Tesoros y ontologías en sistemas de información documental. *El Profesional de la Información*, 20(5), 555-563. <https://doi.org/10.3145/epi.2011.sep.10>

- Codina, L., & Pedraza-Jiménez, R. (2015, marzo 30). Taxonomías y Ontologías: Qué son y cómo se aplican a Medios de Comunicación. Recuperado a partir de <http://www.lluiscodina.com/taxonomias-ontologias/>
- Codina, L., & Pedraza-Jiménez, R. (2017, enero 3). Taxonomías y Ontologías: Qué son y cómo se aplican en medios de comunicación. Recuperado a partir de <https://www.lluiscodina.com/taxonomias-ontologias/>
- Cordón García, J. A., Alonso-Arévalo, J., Gómez Díaz, R., & López Lucas, J. (2012). *Las nuevas fuentes de información: información y búsqueda documental en el contexto de la web 2.0* (2a ed.). Madrid: Ediciones Pirámide.
- Cornella, A. (2010). *Infoxicación: buscando un orden en la información* (2.ª ed.). Barcelona: Zero Factory S.L. Recuperado a partir de http://www.infonomia.com/wp-content/uploads/2014/05/948_infoxicacion.pdf
- Currás, E. (2010). *Ontologies, Taxonomies and Thesauri in System Science and Systematics*. Oxford: Chandox Publishing.
- Daquino, M., Peroni, S., Tomasi, F., & Vitali, F. (2014). Political Roles Ontology (PRoles): Enhancing Archival Authority Records through Semantic Web Technologies. *Procedia Computer Science*, 38, 60-67. <https://doi.org/10.1016/j.procs.2014.10.012>
- Debnath, S., & Giles, C. L. (2005). A learning based model for headline extraction of news articles to find explanatory sentences for events (pp. 189-190). <https://doi.org/10.1145/1088622.1088662>
- Díaz Nosty, B. (2013). *La prensa en el nuevo ecosistema informativo. «¡Que paren las rotativas!»*. *La transición al medio continuo*. Barcelona: Ariel, Fundación Telefónica, Planeta. Recuperado a partir de http://www.fundacion.telefonica.com/es/arte_cultura/publicaciones/detalle/238
- Dutta, B., Nandini, D., & Shahi, G. K. (2015). MOD: Metadata for Ontology Description and Publication. En *DC-2015, São Paulo, Brazil*. Recuperado a partir de

<http://dcevents.dublincore.org/IntConf/dc-2015/paper/view/358>

Escolano Ruiz, F., Cazorla Quevedo, M. Á., Alfonso Galipienso, M. I., Colomina Pardo, O., & Lozano Ortega, M. Á. (2003). *Inteligencia artificial. Modelos, Técnicas y Áreas de Aplicación*. Madrid: Thompson.

Europa Press. (2013, noviembre 23). Pekín impone una «zona de identificación de defensa aérea». *Expansión.com*. Recuperado a partir de http://www.expansion.com/agencia/europa_press/2013/11/23/20131123090449.html

Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I. F., & Couto, F. M. (2013). The AgreementMakerLight Ontology Matching System. En R. Meersman, H. Panetto, T. Dillon, J. Eder, Z. Bellahsene, N. Ritter, ... D. Dou (Eds.), *On the Move to Meaningful Internet Systems: OTM 2013 Conferences* (pp. 527-541). Graz, Austria: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41030-7_38

Fernández-García, N., Arias Fisteus, J., Sánchez, L., & López, G. (2012). IdentityRank: Named Entity Disambiguation in the News Domain. *Expert Syst. Appl.*, 39(10), 9207–9221. <https://doi.org/10.1016/j.eswa.2012.02.084>

Fernández-García, N., Blázquez, J. M., Fisteus, J. A., Sánchez, L., Sintek, M., Bernardi, A., ... Ben-Asher, Z. (2006). NEWS: Bringing Semantic Web Technologies into News Agencies. En I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, ... L. M. Aroyo (Eds.), *The Semantic Web - ISWC 2006* (pp. 778-791). Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/11926078_56

Fernández-García, N., Blázquez, J. M., Sánchez-Fernández, L., & Bernardi, A. (2007). IdentityRank: Named entity disambiguation in the context of the NEWS project (Vol. 4519 LNCS, pp. 640-654). Recuperado a partir de <http://www.scopus.com/inward/record.url?eid=2-s2.0->

34548061897&partnerID=40&md5=09bf835b0eb41d2e206ff59b97ec5ca3

- Fernández-García, N., Fuentes, D., Sánchez, L., & Fisteus, J. A. (2010). The NEWS ontology: Design and applications. *Expert Systems with Applications*, 37(12), 8694-8704. <https://doi.org/10.1016/j.eswa.2010.06.055>
- Fernández-García, N., Sánchez-Fernández, L., Blázquez-del-Toro, J. M., & Villamor-Lugo, J. (2007). The News Ontology for Professional Journalism Applications. En R. Sharman, R. Kishore, & R. Ramesh (Eds.), *Ontologies* (pp. 887-919). Springer US. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-0-387-37022-4_32
- Fidalgo, A. (2004). Sintaxe e semântica das notícias online. Para um jornalismo assente em base de dados. En *Informação e Comunicação Online. Jornalismo Online* (Vol. 1, pp. 49-61). Covilhã, Portugal: Universidade da Beira Interior (UBI). Recuperado a partir de http://www.livroslabcom.ubi.pt/pdfs/20110829-fidalgo_serra_ico1_jornalismo_online.pdf
- Fillmore. (1967). The Case for Case. En E. Bach & R. Harms (Eds.), *Proceeding of the Texas Symposium on Linguistic Universals* (pp. 1-134). New York: Holt, Rinehart, and Winston. Recuperado a partir de <http://eric.ed.gov/?id=ED019631>
- Fondin. (1977). La structure et le vocabulaire de l'analyse documentaire : contribution pour une mise au point. *Documentaliste*, 14(2), 11-16.
- Galfi, R. (2012, septiembre 19). Google News Blog: A newly hatched way to tag your news articles. Recuperado 1 de diciembre de 2014, a partir de <http://googlenewsblog.blogspot.com.es/2012/09/a-newly-hatched-way-to-tag-your-news.html>
- García Gutiérrez, A. (1999). Análisis documental: textos periodísticos escritos. En A. García Gutiérrez (Ed.), *Introducción a la Documentación Informativa y Periodística* (pp. 285-303). Sevilla: Mad.

- García Gutiérrez, A. (2014). Análisis documental de noticias de prensa en sistemas de información factual. *Revista española de Documentación Científica*, 37(2), e046. <https://doi.org/10.3989/redc.2014.2.1094>
- García Jiménez, A. (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. *Anales de Documentación*, 7(0), 79-95.
- García, R., Perdrix, F., & Gil, R. (2006). Ontological Infrastructure for a Semantic Newspaper. En In «*Semantic Web Annotations for Multimedia Workshop, SWAMM 2006*». *15th World Wide Web Conference*.
- García-Hernández, B. (1986). Gramática de casos y disociación intrasubjetiva: dativo profundo y dativo funcional. *Cuadernos de filología clásica*, (20), 331-348.
- García-Hernández, B. G. (1986). Gramática de casos y disociación intrasubjetiva dativo profundo y dativo funcional. *Cuadernos de Filología Clásica*, 20, 231-248. <https://doi.org/>
- García-Marco, F.-J. (2014). La nueva norma de tesauros UNE ISO 25964-1. *Anuario ThinkEPI*, 8(0), 159-163.
- García-Marco, F.-J. (2017). ISO 25964-2: interoperabilidad de los tesauros con otros sistemas de organización del conocimiento en contextos semánticos. *Anuario ThinkEPI*, 11(0), 305-309.
- García-Torres, A., Pareja-Lora, A., & Pradana-López, D. (2008). Reutilización de tesauros: el documentalista frente al reto de la web semántica / Reusing thesauri: documentalists face the semantic web challenge. *El Profesional de la Información*, 17(1), 8-21.
- Gil Leiva, I. (2008). *Manual de indización: teoría y práctica*. Gijón (Asturias): Ediciones Trea S.L.
- Gil Leiva, I. (2012). Aspectos conceituais da indexação. En I. Gil Leiva & M. S. L. Fujita (Eds.), *Política de Indexação* (p. 259). Sao Paulo: Cultura Académica. Recuperado

- a partir de https://www.marilia.unesp.br/Home/Publicacoes/politica-de-indexacao_ebook.pdf
- Gil Pascual, J. A. (2011). *Técnicas e instrumentos para la recogida de información*. Madrid: UNED. UNIVERSIDAD NACIONAL DE EDUCACION A DISTANCIA.
- Gómez, A., Moreno, A., Pazos, J., & Sierra-Alonso, A. (2000). Knowledge maps: An essential technique for conceptualisation. *Data & Knowledge Engineering*, 33(2), 169-190. [https://doi.org/10.1016/S0169-023X\(99\)00050-6](https://doi.org/10.1016/S0169-023X(99)00050-6)
- Gómez-Pérez, A, Fernández, M., & Vicente, A. de. (1996). Towards a Method to Conceptualize Domain Ontologies. En *Proceedings Workshop: Ontological Engineering* (pp. 41-51). Budapest, Rumanía: Facultad de Informática (UPM). Recuperado a partir de <http://oa.upm.es/7228/>
- Gómez-Pérez, Asunción. (2004). Ontology Evaluation. En Steffen Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 251-273). Berlin, Germany: Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-24750-0_13
- Google Inc. (2016). Ayuda para editores: Palabras clave y consultas de búsqueda. Recuperado a partir de <https://support.google.com/news/publisher/answer/68297?hl=es>
- Grenon, P., & Smith, B. (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation*, 4(1), 69–103.
- Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199-220.
- Gruber, T. (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *Int. J. Hum.-Comput. Stud.*, 43(5-6), 907–928. <https://doi.org/10.1006/ijhc.1995.1081>
- Grüninger, M., & Fox, M. S. (1995). Methodology for the Design and Evaluation of

- Ontologies. En *Workshop on Basic Ontological Issues in Knowledge Sharing* (pp. 1-10). Montreal. Recuperado a partir de <http://www.eil.utoronto.ca/enterprise-modelling/papers/gruninger-ijcai95.pdf>
- Guarino, N. (1998). Formal Ontology and Information Systems. En *Proceedings of FOIS'98* (pp. 3–15). Trento, Italy: IOS Press.
- Guerrillot, S. (2006). Use of semantic technologies at Agence France-Presse (AFP). Presentado en Semantic Technology Conference, San José. Recuperado a partir de <http://ceur-ws.org/Vol-194/paper8.pdf>
- Gutiérrez-Cillán, J., & Rodríguez Escudero, I. (2013). Tipos de investigación científica. En *Métodos de investigación social y de la empresa* (pp. 47-66). Madrid: Ediciones Pirámide.
- Haav, H.-M. (2011). A Practical Methodology for Development of a Network of e-Government Domain Ontologies. En T. Skersys, R. Butleris, L. Nemuraite, & R. Suomi (Eds.), *Building the e-World Ecosystem* (pp. 1-13). Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-642-27260-8_1
- Hagey, K. (2012, diciembre 6). Washington Post Plans a Paywall. *Wall Street Journal*. Recuperado a partir de <http://online.wsj.com/article/SB10001424127887324640104578163641549720044.html?mod=e2tw>
- Hartmann, J., Palma, R., & Gómez-Pérez, A. (2009). Ontology Repositories. En S Staab & R. Studer (Eds.), *Handbook on Ontologies* (2a ed., pp. 551-572). Berlin, Germany: Facultad de Informática (UPM). Recuperado a partir de <http://oa.upm.es/6430/>
- Hendler, J., Berners-Lee, T., & Miller, E. (2002). Integrating Applications on the Semantic Web. *Journal of the Institute of Electrical Engineers of Japan*, 22(10), 676-680.
- Hernández-Serrano, M.-J., Renés-Arellano, P., Graham, G., & Greenhill, A. (2017). From

- Prosumer to Prodesigner: Participatory News Consumption. *Comunicar*, 25, 77-88.
<https://doi.org/https://doi.org/10.3916/C50-2017-07>
- Hickson, I., Berjon, R., Faulkner, S., Leithead, T., Navara, E. D., O'Connor, E., & Pfeiffer, S. (Eds.). (2014, octubre 28). HTML5. A vocabulary and associated APIs for HTML and XHTML (W3C Recommendation 28 October 2014). W3C. Recuperado a partir de <https://www.w3.org/TR/html5/document-metadata.html#the-meta-element>
- Hillman, D. (2005, noviembre 7). Using Dublin Core (DCMI Recommended Resource). Recuperado a partir de <http://dublincore.org/documents/usageguide/>
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2009, septiembre 22). OWL 2 Web Ontology Language Primer. W3C Proposed Recommendation 22 September 2009. W3C. Recuperado a partir de <https://www.w3.org/TR/2009/PR-owl2-primer-20090922/all.pdf>
- Hitzler, P., & Parsia, B. (2009). Ontologies and Rules. En Steffen Staab & R. Studer (Eds.), *Handbook on Ontologies* (2a ed., pp. 111-132). Berlín: Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-92673-3_5
- Hlomani, H. (2014, septiembre 2). *Multidimensional Data-driven Ontology Evaluation*. University of Guelph, Guelph, Ontario, Canada. Recuperado a partir de <http://hdl.handle.net/10214/8380>
- Hobbs, J. R., & Pan, F. (Eds.). (2006, septiembre 27). Time Ontology in OWL. W3C Working Draft 27 September 2006. W3C. Recuperado a partir de <http://www.w3.org/TR/owl-time/>
- Horridge, M. A. (2011). *Practical Guide to Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.3*. Manchester: University of Manchester. Recuperado a partir de <http://mowl->

- power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf
- Huntington, S. P. (2001). *El choque de civilizaciones y la recuperación del orden mundial*. (J. P. Tosaus Abadía, Trad.). Buenos Aires, Argentina: Paidós.
- Ingersoll, G. S., Morton, T. S., & Farris, A. L. (2013). *Taming text. How to find, organize and manipulate it*. Shelter Island, NY: Manning Publications Co.
- Iqbal, R., Murad, M. A. A., Mustapha, A., Sharef, & Mohd, N. (2013). An Analysis of Ontology Engineering Methodologies: A Literature Review. *Research Journal of Applied Sciences, Engineering and Technology*, 6(16), 2993-3000.
- Jiménez Cano, R., & Abad Liñán, J. M. (2015, mayo 13). 'The New York Times', 'The Guardian' y la BBC publican directamente en Facebook desde hoy. *El País*. Recuperado a partir de http://tecnologia.elpais.com/tecnologia/2015/05/13/actualidad/1431490102_473389.html
- Jokela, S., Turpeinen, M., Kurki, T., Savia, E., & Sulonen, R. (2001). The role of structured content in a personalized news service. *Acta Polytechnica Scandinavica Mathematics and Computing Series*, (114), XX-XXI.
- Kallipolitis, L., Karpis, V., & Karali, I. (2007). World news finder: How we cope without the Semantic Web. En V. Devedzic (Ed.), *Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications* (pp. 616-621). Innsbruck, Austria: Acta Press Anaheim.
- Kallipolitis, L., Karpis, V., & Karali, I. (2012). Semantic search in the World News domain using automatically extracted metadata files. *Knowledge-Based Systems*, 27, 38-50. <https://doi.org/10.1016/j.knosys.2011.12.007>
- Kasper, W., Steffen, J., & Zhang, Y. (2008a). News Annotations for Navigation by Semantic Similarity. En A. R. Dengel, K. Berns, T. M. Breuel, F. Bomarius, & T. R. Roth-Berghofer (Eds.), *KI 2008: Advances in Artificial Intelligence* (pp. 233-240).

- Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-85845-4_29
- Kasper, W., Steffen, J., & Zhang, Y. (2008b). Using Semantics for News Navigation. En *2008 IEEE International Conference on Semantic Computing* (pp. 261-267). <https://doi.org/10.1109/ICSC.2008.42>
- Keyser, P. de. (2012). *Indexing from thesauri to the semantic web*. Oxford: Chandox Publishing.
- Kim, S., Iglesias-Sucasas, M., & Viollier, V. (2013). The FAO Geopolitical Ontology: A Reference for Country-Based Information. *Journal of Agricultural & Food Information*, 14(1), 50-65. <https://doi.org/10.1080/10496505.2013.747193>
- Kodama, M., Ozono, T., Shintani, T., & Aosaki, Y. (2008). Realizing a news value markup language for news management systems using newsML. En F. Xhafa & L. Barolli (Eds.), *2nd International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 249-255). Los Alamitos: Ieee Computer Soc. <https://doi.org/10.1109/CISIS.2008.70>
- Lamarca Lapuente, M. J. (2013a, agosto 12). *Hipertexto, el nuevo concepto de documento en la cultura de la imagen: Indización automática* (tesis). Universidad Complutense de Madrid, Madrid. Recuperado a partir de http://www.hipertexto.info/documentos/indiz_automat.htm
- Lamarca Lapuente, M. J. (2013b, agosto 12). *Hipertexto, el nuevo concepto de documento en la cultura de la imagen: Metadatos en HTML* (tesis). Universidad Complutense de Madrid, Madrid. Recuperado a partir de http://www.hipertexto.info/documentos/meta_html.htm
- Lambe, P. (2007). *Organising knowledge: Taxonomies, knowledge and organisational effectiveness*. Oxford: Chandox Publishing.
- Lancaster, F. W. (2002). *El control del vocabulario en la recuperación de información*. (A.

- De la Cueva, Trad.) (2a ed.). Valencia: Universidad de Valencia.
- Lancaster, F. W. (2004). *Indexação e resumos: Teoria e prática*. (A. A. B. de Lemos, Trad.) (2a ed.). Brasília, DF: Briquet de Lemos / Livros.
- Liddy, E. D., McVearry, K. A., Paik, W., Yu, E., & McKenna, M. (1993). Development, Implementation and Testing of a Discourse Model for Newspaper Texts. En *Proceedings of the Workshop on Human Language Technology* (pp. 159–164). Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.3115/1075671.1075707>
- Litton, G. (1971). *La Documentación: Breviario del bibliotecario*. Argentina: Bowker.
- Maculan, B. C. M. dos S. (2014). *Taxonomia Facetada e Navegacional: Um Mecanismo de Recuperação*. Curitiba: Appris. Recuperado a partir de <http://www.editoraappris.com.br/produto/4339991/Taxonomia-Facetada-e-Navegacional-Um-Mecanismo-de-Recuperacao>
- Maculan, B. C. M. dos S., & Lima, G. (2011). Taxonomia facetada navegacional: agregando valor às informações disponibilizadas em bibliotecas digitais de teses e dissertações. En *Enancib* (Vol. 12). Brasília - DF, Brazil: Universidade de Brasília, Faculdade de Ciência da Informação. <https://doi.org/2011-0000666-00038>
- Maidel, V., Shoval, P., Shapira, B., & Taieb-Maimon, M. (2010). Ontological content-based filtering for personalised newspapers: A method and its evaluation. *Online Information Review*, 34(5), 729-756. <https://doi.org/10.1108/14684521011084591>
- Mannens, E., Coppens, S., De Pessemier, T., Dacquin, H., Van Deursen, D., De Sutter, R., & Van de Walle, R. (2013). Automatic news recommendations via aggregated profiling. *Multimedia Tools and Applications*, 63(2), 407-425. <https://doi.org/10.1007/s11042-011-0844-8>
- Martín Muñoz, D. (2008). Mapeo y Fusión. Recuperado 27 de mayo de 2016, a partir de <http://ontologiasdemetadatos.50webs.com/mapeoyfusion.html>

- Martínez Mediano, C. (2014). *Técnicas e instrumentos de recogida y análisis de datos*. Madrid: UNED. UNIVERSIDAD NACIONAL DE EDUCACION A DISTANCIA.
- Martínez Méndez, F. J. (2002). *Propuesta y desarrollo de un modelo para la evaluación de la recuperación de información en internet*. Universidad de Murcia, Murcia. Recuperado a partir de <http://www.tdx.cat/handle/10803/10904>
- Martínez Méndez, F. J. (2004). *Recuperación de información: modelos, sistemas y evaluación*. Murcia: JMC Kiosko Ediciones. Recuperado a partir de <http://bibliotecadigital.org/jspui/handle/001/227>
- Martínez-Béjar, R., Benjamins, V. R., & Martín-Rubio, F. (1997). Designing operators for constructing domain knowledge ontologies. En *Knowledge Acquisition, Modeling and Management* (pp. 159-173). Sant Feliu de Guixols, Cataluña, España: Springer, Berlin, Heidelberg. <https://doi.org/10.1007/BFb0026784>
- Martínez-Fernández, J. L., García-Serrano, A., Martínez, P., & Villena, J. (2004). Automatic Keyword Extraction for News Finder. En A. Nürnberger & M. Detyniecki (Eds.), *Adaptive Multimedia Retrieval* (pp. 99-119). Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-25981-7_7
- Martínez-Béjar, R., Ibañez-Cruz, F., Compton, P., & Cao, T. M. (2001). An easy-maintenance, reusable approach for building knowledge-based systems: application to landscape assessment. *Expert Systems with Applications*, 20(2), 153-162. [https://doi.org/10.1016/S0957-4174\(00\)00055-5](https://doi.org/10.1016/S0957-4174(00)00055-5)
- McGee, M. (2012, septiembre 19). Google Announces News Keywords Meta Tag For Publishers. Recuperado 27 de abril de 2015, a partir de <http://searchengineland.com/google-announces-news-keywords-metatag-133759>
- McGinnis, J., Wilton, P., Harman, P., & O'Donovan, J. (2011). SNaP Ontologies. Simple News and Press Ontologies. Recuperado a partir de <http://data.press.net/ontology/>

- Méndez Rodríguez, E., & Greenberg, J. (2012). Datos enlazados para vocabularios abiertos y marco general de HIVE. *El profesional de la información*, 21(3), 236-244.
- Mendonça, F. M., & Almeida, M. B. (2014). Princípios metodológicos para desenvolvimento de ontologias: análise das práticas correntes e proposição de melhorias. *BENANCIB*. Recuperado a partir de <http://hdl.handle.net/123456789/2602>
- Mesquita, N. C., Moisés, J. Á., & Eldasi, G. (2015). O projeto Corrupteca: uma ferramenta analítica para a análise do fenômeno da corrupção. *Media & Jornalismo*, 14(26), 65-76.
- Mizoguchi, R., & Kozaki, K. (2009). Ontology Engineering Environments. En Steffen Staab & R. Studer (Eds.), *Handbook on Ontologies* (2a ed., pp. 315-336). Berlin, Germany: Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-92673-3_14
- Moens, M.-F. (2000). *Automatic indexing and abstracting of document text*. Norwell, Massachusetts, USA: Kluwer Academic Publisher.
- Morales-del-Castillo, J. M. (2008). *Modelo de servicio semántico-difuso de difusión selectiva de información (DSI) para bibliotecas digitales*. Universidad de Granada, Granada. Recuperado a partir de <http://www.tdx.cat/handle/10803/16603>
- Moreira, S., Batista, D., Carvalho, P., Couto, F. M., & Silva, M. J. (2011). POWER - Politics Ontology for Web Entity Retrieval. En C. Salinesi & O. Pastor (Eds.), *Advanced Information Systems Engineering Workshops* (pp. 489-500). London, UK: Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-642-22056-2_51
- Moreira, S., Batista, D. S., Carvalho, P., Couto, F. M., & Silva, M. J. (2013). Tracking politics with POWER. *Program: electronic library and information systems*, 47(2), 120-135. <https://doi.org/10.1108/00330331311313708>

- Moreno Sandoval, A., & Campillos Llanos, L. (2015). Combinación de estrategias léxicas y estadísticas para el reconocimiento automático de términos: su aplicación a un corpus de medicina. *LEA: Lingüística española actual*, 37(2), 175-199.
- Motz, R., Badell, C., Barrosa, M., Sum, R., Díaz, G., & Castro, M. (2009). LooKIng4LO: Sistema Informático para la Extracción Automática de Objetos de Aprendizaje: Caso de Estudio. *Revista Iberoamericana De Tecnologías Del Aprendizaje - IEEE-RITA*, 4(3), 223-229.
- Moyano Grimaldo, W. A. (2017). Adaptabilidad de la Clasificación Decimal Dewey para la organización de contenidos: de los estantes a la Web/Adapting the Dewey Decimal Classification to Content Organization: From Shelves to the Web. *Revista Interamericana de Bibliotecología; Medellín*, 40(1), 27-34. <https://doi.org/http://dx.doi.org/10.17533/udea.rib.v40n1a03>
- Narukawa, C. M., Gil Leiva, I., & Fujita, M. S. L. (2009). Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia DeCS na área de Odontologia. *Informação & Sociedade: Estudos*, 19(2), 99-118.
- Neches, R., Fikes, R., Finin, T., Gruber, T., Ramesh, P., Senator, T., & Swartout, W. R. (1991). Enabling Technology for Knowledge Sharing. *AI Magazine*, 12(3), 36-56.
- Nicolino, M. E. V. P., & Ferneda, E. (2014). Um Método para a Utilização de Ontologias na Indexação Automática. *Informação & Tecnologia*, 1(2), 13-33.
- Nies, T. de, D'heer, E., Coppens, S., Van Deursen, D., Mannens, E., & Van de Walle, R. (2012). Bringing newsworthiness into the 21st century. En *Web of Linked Entities, Workshop proceedings* (pp. 106-117). Recuperado a partir de <http://ceur-ws.org/Vol-906/paper11.pdf>
- NISO. (2005). *ANSI/NISO Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies (version of record for*

- 2010 reaffirmation of 2005 edition). Recuperado a partir de http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf
- Nowak, B. (2009, julio 8). The Semantic Web - Not a piece of cake... Recuperado 5 de octubre de 2016, a partir de <http://bnode.org/blog/2009/07/08/the-semantic-web-not-a-piece-of-cake>
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Recuperado a partir de http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html
- Ochoa Hernández, J. L. (2011). *Desarrollo de una metodología para la construcción automática de ontologías en español a partir de texto libre / José Luis Ochoa Hernández; directores, Rafael Valencia García, Rodrigo Martínez Béjar*. (info:eu-repo/semantics/doctoralThesis). Universidad de Murcia. Recuperado a partir de <https://digitum.um.es/xmlui/handle/10201/32892>
- Ogden, C. K., & Richards, I. A. (1923). *The meaning of the meaning, a study of the influence of the language upon thought and of the science of symbolism*. New York, USA: Harvest Book. Recuperado a partir de <http://s-f-walker.org.uk/pubsebooks/pdfs/ogden-richards-meaning-all.pdf>
- Oliveira, L. H. M. de. (2009a). *e-Termos, ambiente colaborativo Web de Gestão Terminológica*. Embrapa Informática Agropecuária (CNPTIA), Universidade de São Paulo (USP Campus de São Carlos, SP) e Universidade Federal de São Carlos (UFSCar). Recuperado a partir de <https://www.etermos.cnptia.embrapa.br/index.php>
- Oliveira, L. H. M. de. (2009b, septiembre 22). *e-Termos: Um ambiente colaborativo web de gestão terminológica* (text). Universidade de São Paulo, São Paulo. Recuperado

- a partir de <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-19012010-150638/>
- Ortiz, A. (2011). Polionto: Ontology reuse with Automatic Text Extraction from Political Documents. En A. A. Sousa & E. Oliveira (Eds.), *Proceedings of the 6th doctoral symposium in informatics engineering* (pp. 309-320). Porto (Portugal): Faculdade de Engenharia, Universidade do Porto.
- Otlet, P. (1934). *El tratado de documentacion 2º edición*. (M. D. Ayuso García, Trad.) (2.^a ed.). Murcia: Editum. Ediciones de la Universidad de Murcia.
- Paepen, B. E. (2002). Omnipaper: Bringing electronic news publishing to a next level using XML and Artificial Intelligence. En J. A. Carvalho, A. Hübler, & A. A. Baptista (Eds.), *Proceedings of the 6th International ICCO/IFIP Conference on Electronic Publishing*. Karlovy Vary, Czech Republic: VWF Berlin. Recuperado a partir de <http://elpub.scix.net/cgi-bin/works/Show?02-29>
- Pan, J. Z. (2009). Resource Description Framework. En Stefen Staab & R. (eds. . Studer (Eds.), *Handbook on Ontologies* (2a ed., pp. 71-90). Berlín: Springer.
- Pastor-Sánchez, J. A. (2011). *Tecnologías de la web semántica* (Edición: 1). Barcelona: Editorial UOC, S.L.
- Pastor-Sánchez, J.-A. (2017). Vocabularios para la descripción de conjuntos de datos en la Web. *Anuario ThinkEPI*, 11(0), 278-283. <https://doi.org/10.3145/thinkepi.2017.54>
- Pedraza-Jiménez, R., Codina, L., & Rovira, C. (2007). Web semántica y ontologías en el procesamiento de la información documental. *El Profesional de la Información*, 16(6), 569-578. <https://doi.org/10.3145/epi.2007.nov.04>
- Peis Redondo, E., Herrera-Viedma, E., Hassan-Montero, Y., & Herrera, J. C. (2003a). Análisis de la web semántica: estado actual y requisitos futuros. *El Profesional de la Información*, 12(5), 368-376.
- Peis Redondo, E., Herrera-Viedma, E., Hassan-Montero, Y., & Herrera, J. C. (2003b).

- Ontologías, metadatos y agentes: recuperación «semántica» de la información. En *Actas de las II Jornadas de Tratamiento y Recuperación de la Información (JOTRI)* (pp. 157-165). Madrid. Recuperado a partir de <http://hdl.handle.net/10481/1206>
- Pellegrini, T. (2012). Integrating linked data into the content value chain - A review of news-related standards, methodologies and licensing requirements. En *ACM International Conference Proceeding Series* (pp. 94-102).
- Pellegrini, Tassilo. (2012). Semantic metadata in the news production process: achievements and challenges. En *Proceedings of the 16th International Academic MindTrek Conference 2012* (pp. 125-133). Tampere; Finland: ACM Press. <https://doi.org/10.1145/2393132.2393158>
- Pena, R. A. P. (2012, enero). *Suporte semântico à publicação de conteúdo jornalístico na Web* (Tesis de Máster). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro. Recuperado a partir de <http://doi.org/10.17771/PUCRio.acad.20235>
- Pereira, T., & Baptista, A. A. (2003). Omnipaper: descrição de recursos de notícias digitais em RDF. En J. C. Ramalho, Henriques Pedro Rangel, G. R. Librelotto, & G. V. Arnold (Eds.), *XML, aplicações e tecnologias associadas*. Braga, Portugal: Universidade do Minho. Recuperado a partir de <http://repositorium.sdum.uminho.pt/handle/1822/283>
- Pereira, T., & Baptista, A. A. (2004). Incorporating a semantically enriched navigation layer onto an RDF metadatabase. En J. Engelen, S. M. S. Costa, & A. C. S. Moreira (Eds.), *Building digital bridges: linking cultures, commerce and science*. Brasília. Recuperado a partir de <http://repositorium.sdum.uminho.pt/handle/1822/604>
- Pereira, T. S. M., & Baptista, A. A. (2003). The omnipaper metadata RDF/XML prototype implementation. En S. M. SouzaCosta, J. A. Carvalho, A. A. Baptista, & A. C. S. Moreira (Eds.), *Information to knowledge: proceedings* (pp. 261-268). Guimaraes:

- Universidade Minho Dept Sistemas Informacao. Recuperado a partir de <http://repositorium.sdum.uminho.pt/handle/1822/171>
- Pereira, Teresa Susana Mendes, & Baptista, A. A. (2009). The instantiation of OmniPaper RDF prototype in the context of scientific publications. *Electronic Library*, 27(5), 767-778. <https://doi.org/10.1108/02640470910998506>
- Peroni, S., Shotton, D., & Vitali, F. (2012). Making Ontology Documentation with LODÉ. En *Proceedings of the I-SEMANTICS 2012 Posters & Demonstrations Track* (pp. 63-67). Graz, Austria. Recuperado a partir de <https://pdfs.semanticscholar.org/df05/e1df6363a78739f3c1db0f8c562ccc24e209.pdf>
- Quandt, T. (2008). (NO) NEWS ON THE WORLD WIDE WEB? A comparative content analysis of online news in Europe and the United States. *Journalism Studies*, 9(5), 717-738. <https://doi.org/10.1080/14616700802207664>
- Raber, D. (2003). *The problem of information: An introduction to Information Science*. United States of America: Scarecrow Press, Inc.
- RAE. (2014a). Apartheid. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=36AsHJ3>
- RAE. (2014b). Cosa. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=B3yTydM>
- RAE. (2014c). Evento. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=H9JpZQS>
- RAE. (2014d). Explotar¹. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=HKDxVB1|HKE4K7Q>
- RAE. (2014e). Fecha. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=HhpCnjk>
- RAE. (2014f). Herencia. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir

- de <http://dle.rae.es/?id=KCSxJPx>
- RAE. (2014g). Instrumento. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=LoJdDcs>
- RAE. (2014h). Política. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=Ta2HMYR>
- RAE. (2014i). Polución. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=TbYjNMb>
- RAE. (2014j). Sitio. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=Y2qwiA7>
- RAE. (2014k). Sujeto. *Diccionario de la lengua española* (23.^a ed.). Recuperado a partir de <http://dle.rae.es/?id=YgC4A98>
- Raimond, Y., & Abdallah, S. (2007, octubre 25). The Event Ontology [Version 1.0]. Recuperado a partir de <http://motools.sourceforge.net/event/event.html>
- Río Cabrerizo, R. del. (2014). *Periodismo económico y financiero*. Madrid: Síntesis.
- Rocha, L. B. L. da. (2012). *Ontologia de notícias: Um modelo para classificação do conteúdo dos jornais on-line brasileiros, segundo a lógica da Web Semântica* (Tesis de Máster). Universidade do Estado do Rio de Janeiro, Rio de Janeiro.
- Rocha, L. B. L. da, Lessa, W., & Monat, A. (2009). Classificação do conteúdo dos jornais online, segundo a lógica da Web Semântica. En *Conference on Human Factors in Computing Systems*. Boston, Massachusetts, USA. Recuperado a partir de http://www.estacazero.com/publicacoes/ixdsa11_webjornais_artigo_submetido.pdf
- Rodríguez Alberich, G. (2014). Dirae.es: Qué es. Recuperado 16 de agosto de 2016, a partir de <http://dirae.es/about>
- Rodríguez Mateos, D., & Pérez Lorenzo, B. (2011). Análisis documental de contenidos audiovisuales. En *Documentación audiovisual: Nuevas tendencias en el entorno digital* (p. 231). Madrid: Síntesis.

- Rodríguez Perojo, K., & Ronda León, R. (2005). Web semántica: un nuevo enfoque para la organización y recuperación de información en el web. *Revista Cubana de Información en Ciencias de la Salud (Acimed)*, 13(6). Recuperado a partir de http://bvs.sld.cu/revistas/aci/vol13_6_05/aci030605.htm
- Rodríguez Yunta, L. (2006). Una lectura del Diccionario Panhispánico de dudas desde el punto de vista de los lenguajes documentales. *Revista General de Información y Documentación*, 16(1), 197-206.
- Romero Llop, R. (2007). *Especificación OWL de una ontología para teleeducación en la web semántica*. Universidad Politécnica de Valencia, Valencia. Recuperado a partir de <http://hdl.handle.net/10251/1828>
- Roussey, C., Pinet, F., Kang, M. A., & Corcho, O. (2011). An Introduction to Ontologies and Ontology Engineering. En *Ontologies in Urban Development Projects* (pp. 9-38). Springer London. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-0-85729-724-2_2
- Rubio Lacoba, M. (2012). Nuevas destrezas documentales para periodistas: el vocabulario colaborativo del diario El País. *Trípodos*, 31, 65-78.
- Rubio Lacoba, M., & Blanco García, J. C. (2010). Mejor que un buscador, un encontrador. *Documentación de las Ciencias de la Información*, 33(0), 273-287. <https://doi.org/>
- Ruíz-Martínez, J. M. (2011). *Metodología para la población automática de ontologías. Aplicación en los dominios de medicina y turismo*. Universidad de Murcia, Murcia.
- Ruiz-Martínez, J. M., Baños-Moreno, M.-J., & Martínez-Béjar, R. (2014). Nomenclatura Unesco: evolución, alcance y reutilización en clave ontológica para la descripción de perfiles científicos. *El Profesional de la Información*, 23(4), 383-392.
- Saias, J., & Quaresma, P. (2007). The Senso Question Answering Approach to Portuguese QA@CLEF-2007. En A. Nardi, C. Peters, & N. Ferro (Eds.), *Working Notes for CLEF 2007 Workshop co-located with the 11th European Conference on Digital*

- Libraries (ECDL 2007)*, Budapest, Hungary, September 19-21, 2007 (Vol. 1173). CEUR-WS.org. Recuperado a partir de <http://ceur-ws.org/Vol-1173/CLEF2007wn-QACLEF-SaiasEt2007.pdf>
- Saias, J., & Quaresma, P. (2008). The Senso Question Answering System at QA@CLEF 2008. En C. Peters & N. Ferro (Eds.), *Working Notes for CLEF 2008 Workshop co-located with the 12th European Conference on Digital Libraries (ECDL 2008)*, Aarhus, Denmark, September 17-19, 2008 (Vol. 1174). CEUR-WS.org. Recuperado a partir de <http://ceur-ws.org/Vol-1174/CLEF2008wn-QACLEF-SaiasEt2008.pdf>
- Saleh, L. M. B., & Al-Khalifa, H. S. (2009). AraTation: An Arabic semantic annotation tool (pp. 447-451). <https://doi.org/10.1145/1806338.1806421>
- Salton, G. (1975). *Dynamic information and library processing*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Salton, G., & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill.
- Sanchez-Cuadrado, S., Morato-Lara, J., Palacios-Madrid, V., Llorens-Morillo, J., & Moreiro-González, J.-A. (2012). De repente, ¿todos hablamos de ontologías? *El profesional de la información*, 16(6), 562-567.
- Sánchez-Fernández, L., Bernardi, A., & Fuentes, M. (2005). An experience with Semantic Web technologies in the news domain. En *Integration of Knowledge, Semantics and Digital Media Technology, 2005. EWIMT 2005. The 2nd European Workshop on the (Ref. No. 2005/11099)* (pp. 455-459). London.
- Santos, L. C. de M. dos, & Bräscher, M. (2015). Recuperação da informação em acervos digitais de jornais impressos: proposta para modelagem de ontologia no domínio do futebol. En *XVI Encontro Nacional de Pesquisa em Ciência da Informação*. Joao Pessoa, PB. Recuperado a partir de <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2790>

- Santos, M. L. F. N. dos. (2007). *Organização do conhecimento e representação de assuntos*. Lisboa: Biblioteca Nacional.
- Saussure, F. (1916). *Curso de lingüística general*. (A. Alonso, Trad.) (24.^a ed.). Buenos Aires, Argentina: Losada.
- Scharffe, F., Zamazal, O., & Fensel, D. (2014). Ontology alignment design patterns. *Knowledge and Information Systems*, 40(1), 1-28. <https://doi.org/10.1007/s10115-013-0633-y>
- Scherp, A., Franz, T., Saathoff, C., & Staab, S. (2009). F—a Model of Events Based on the Foundational Ontology Dolce+DnS Ultralight. En *Proceedings of the Fifth International Conference on Knowledge Capture* (pp. 137–144). New York, NY, USA: ACM. <https://doi.org/10.1145/1597735.1597760>
- Schranz, M. (2005). OmniPaper: Towards a Universal Standard Model for Efficient Information Retrieval. En M. Dobрева, J. Engelen, & Peeters Publishing Leuven (Eds.), *Challenges for the Digital Content Chain: Proceedings of the 9th ICCCE International Conference on Electronic*. Leuven-Heverlee (Belgium): Katholieke Universiteit Leuven. Recuperado a partir de <http://elpub.scix.net/cgi-bin/works/Show?227elpub2005>
- Schranz, M., Dustdar, S., & Platzer, C. (2005). Building an Integrated Pan-European News Distribution Network. En L. M. Camarinha-Matos, H. Afsarmanesh, & A. Ortiz (Eds.), *Collaborative Networks and Their Breeding Environments* (pp. 587-596). Springer US. Recuperado a partir de http://link.springer.com/chapter/10.1007/0-387-29360-4_62
- Schreiber, G., & Raimond, Y. (Eds.). (2014, junio 24). RDF 1.1 Primer (W3C Working Group Note 24 June 2014). W3C. Recuperado a partir de <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>
- Schwartz, C. (2001). *Sorting out the web. Approaches to Subject Access*. Wetsport: Ablex

Publishing.

- Shapira, B., Shoval, P., Tractinsky, N., & Meyer, J. (2009). ePaper: A personalized mobile newspaper. *Journal of the American Society for Information Science and Technology*, 60(11), 2333–2346. <https://doi.org/10.1002/asi.21172>
- Shaw, R., Troncy, R., & Hardman, L. (2009). LOD: Linking Open Descriptions of Events. En *Proceedings of the 4th Asian Conference on The Semantic Web* (pp. 153–167). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-10871-6_11
- Shayji, S. A., Nahla, D., Zant, E., Kadhi, E., & Wong, P. Z. (2011). *Building Ontology for the Political Domain*.
- Shotton, D., & Peroni, S. (2013, diciembre 18). PRO, the Publishing Roles Ontology (versión 1.5.3). Recuperado a partir de <http://purl.org/spar/pro>
- Shoval, P., Maidel, V., & Shapira, B. (2008). An Ontology- Content-based Filtering Method. *International Journal ITA*, 15(4), 303-314.
- Shvaiko, P., & Euzenat, J. (2005). A Survey of Schema-Based Matching Approaches. En S. Spaccapietra (Ed.), *Journal on Data Semantics IV* (pp. 146-171). Springer Berlin Heidelberg. https://doi.org/10.1007/11603412_5
- Silva, D. L. da. (2008, junio 27). *Uma proposta metodológica para construção de ontologias: uma perspectiva interdisciplinar entre as Ciências da Informação e da Computação* (Dissertação de Mestrado). Universidade Federal de Minas Gerais, Belo Horizonte. Recuperado a partir de <http://hdl.handle.net/1843/ECID-7NRQZ2>
- Silva, D. L. da, Souza, R. R., & Almeida, M. B. (2008). Ontologies and controlled vocabularies: comparison of methodologies for construction. *Ciência da Informação*, 37(3), 60–75.
- Silveira, D. T., & Córdova, F. P. (2009). A pesquisa científica. En *Métodos de pesquisa* (pp. 31-42). Porto Alegre, Brasil: Editora da UFRGS. Recuperado a partir de

<http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>

- Simperl, E., Mochol, M., & Bürger, T. (2010). Achieving maturity: the state of practice in ontology engineering in 2009. *International Journal of Computer Science and Applications*, 7(1), 45-65.
- Slype, G. van. (1991). *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. (P. Hípola & F. Moya Anegón, Trads.). Madrid: Fundación Germán Sánchez Ruipérez; Pirámide. Recuperado a partir de <http://eprints.rclis.org/18372/>
- Soler Monreal, C., & Gil Leiva, I. (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías. *Revista Interamericana de Bibliotecología*, 33(2), 361-377.
- Sosa, E. (1997). Procesamiento del lenguaje natural: revisión del estado actual, bases teóricas y aplicaciones (Parte I). *El Profesional de la Información*, 6(1-2), 26-29.
- Souza Brandão Guaraldo, T. de, & Almeida Júnior, O. F. de. (2010). Leitura, informação e conhecimento: notas sobre a leitura de jornal. En *Gestão, mediação e uso da informação* (pp. 191-209). São Paulo: Editora UNESP; Cultura Acadêmica. Recuperado a partir de <http://books.scielo.org/id/j4gkh/10>
- Stevens, R., Egaña Aranguren, M., Wolstencroft, K., Sattler, U., Drummond, N., Horridge, M., & Rector, A. (2007). Using OWL to model biological knowledge. *International Journal of Human-Computer Studies*, 65(7), 583-594. <https://doi.org/10.1016/j.ijhcs.2007.03.006>
- Stuart, D. (2016). *Practical Ontologies for Information Professionals*. London, UK: Facet Publishing.
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1&€“2), 161-197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)

- Tarling, J., & McGinnis, J. (2013, mayo 1). Storyline Ontology. Recuperado a partir de <http://www.bbc.co.uk/ontologies/storyline>
- Taylor, A. G. (1999). *The Organization of Information*. Englewood: Libraries Unlimited.
- Tenenboim, L., Shapira, B., & Shoval, P. (2008). Ontology-based classification of news in an electronic newspaper (pp. 89-97). Presentado en INFOS 2008: Intelligent Information and Engineering Systems, Varna, Bulgaria. Recuperado a partir de <http://sci-gems.math.bas.bg/jspui/bitstream/10525/1035/1/IBS-02-p12.pdf>
- Tennis, J. T., & Calzada-Prado, J. (2007). Ontologies and the Semantic Web: Problems and Perspectives for LIS Professionals. *IBERSID: Journal of Information and Documentation Systems*, 1, 303-311.
- Toffler, A. (1973). *El shock del futuro*. (J. Ferrer Aleu, Trad.). Barcelona: Plaza & Janés. Recuperado a partir de <https://pciucr.files.wordpress.com/2011/03/toffler-alvin-el-shock-del-futuro.pdf>
- Toffler, A. (1980). *The third wave: The classic study of tomorrow*. New York, NY, USA: Bantam.
- Troncy, R. (2008). Bringing the IPTC News Architecture into the Semantic Web. En A. Sheth, S. Staab, M. Paolucci, D. Maynard, T. Finin, & T. Krishnaprasad (Eds.), *Semantic Web - Iswc 2008* (Vol. 5318, pp. 483-498). Berlin: Springer-Verlag Berlin. <https://doi.org/10.1007/978-3-540-88564-1-31>
- Turpo Chaparro, J. E., & Medina, G. E. (2013). Producción intelectual y visibilidad científica. *Apuntes Universitarios*, 3(2), 9-18.
- Uschold, M., & Grüninger, M. (1996). Ontologies: Principles, methods and applications. *KNOWLEDGE ENGINEERING REVIEW*, 11, 93-136.
- Uschold, M., & King, M. (1995). Towards a Methodology for Building Ontologies. En *Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal. Recuperado a partir de

- http://www1.cs.unicam.it/insegnamenti/reti_2008/Readings/Uschold95.pdf
- Vandamme, S., Deleu, J., Wauters, T., Vermeulen, B., & De Turck, F. (2009). CROEQS: Contemporaneous Role Ontology-based Expanded Query Search - Implementation and Evaluation. En *International Conference on Communication Software and Networks, 2009. ICCSN '09* (pp. 448-452). Macau, China. <https://doi.org/10.1109/ICCSN.2009.159>
- Vrandečić, D. (2009). Ontology Evaluation. En Steffen Staab & R. Studer (Eds.), *Handbook on Ontologies* (2a ed., pp. 293-313). Berlin, Germany: Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-92673-3_13
- Wang, W., & Zhao, D. (2012). Ontology-Based Event Modeling for Semantic Understanding of Chinese News Story. En M. Zhou, G. Zhou, D. Zhao, Q. Liu, & L. Zou (Eds.), *Natural Language Processing and Chinese Computing* (pp. 58-68). Springer Berlin Heidelberg. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-642-34456-5_6
- Wikipedia. (2015, noviembre 27). Red eléctrica. En *Wikipedia, la enciclopedia libre*. Recuperado a partir de https://es.wikipedia.org/w/index.php?title=Red_el%C3%A9ctrica&oldid=87300959
- Wikipedia. (2016, noviembre 29). Central nuclear de Kudankulam. En *Wikipedia, la enciclopedia libre*. Recuperado a partir de https://es.wikipedia.org/w/index.php?title=Central_nuclear_de_Kudankulam&oldid=95300581
- Wikipedia. (2017, enero 21). Departamento de Defensa de los Estados Unidos. En *Wikipedia, la enciclopedia libre*. Recuperado a partir de https://es.wikipedia.org/w/index.php?title=Departamento_de_Defensa_de_los_Estados_Unidos&oldid=96350708
- Yaginuma, T., Pereira, T., Ariza Ávila, C. E., & Baptista, A. A. (2004). Implementation of

- metadata for OmniPaper RDF prototype. En *s.n.* Tsukuba, Ibaraki, Japan. Recuperado a partir de <http://repositorium.sdum.uminho.pt/handle/1822/606>
- Yaginuma, T., Pereira, T., & Baptista, A. A. (2003a). Design of metadata elements for digital news articles in the omnipaper project. En S. M. de Souza Costa, J. A. Carvalho, A. A. Baptista, & A. C. Santos Moreira (Eds.), *From information to knowledge: 7th ICC/IFIP International Conference on Electronic Publishing* (pp. 132-139). Minho, Portugal: Universidade do Minho. Recuperado a partir de <http://repositorium.sdum.uminho.pt/handle/1822/170>
- Yaginuma, T., Pereira, T., & Baptista, A. A. (2003b). Metadata elements for digital news resource description. En *Proceedings CLME'2003 - 3º Congresso Luso-Moçambicano de Engenharia* (pp. 1317-1326). Maputo. Recuperado a partir de <http://repositorium.sdum.uminho.pt/handle/1822/279>
- Yedid, N. (2016). Análisis práctico de folksonomías: el caso de los gestores bibliográficos sociales. *Información, Cultura y Sociedad*, 35, 31-64.
- Yves, R., & Abdallah, S. (2007). The Event Ontology. Recuperado a partir de <http://motools.sourceforge.net/event/event.122.html#event>

*

**

ANEXOS

(...) search is only as good as the content backing that search. If the words and phrases that your users are looking for aren't in your index, then you won't be able to return a relevant result (Ingersoll et al., 2013, p. 12)

A. DESCRIPCIÓN DE VOCABULARIOS MAPEADOS

Se describen los vocabularios seleccionados, siguiendo esta estructura: introducción al vocabulario; clases, propiedades tipo objeto y propiedades tipo dato; vocabularios reutilizados. Algunas ontologías tienen propiedades de anotación. En términos generales, no se han descrito, salvo que se consideren de interés. Por lo que respecta a los nombres de los elementos es *CamelBack*, como se describe en 2.2.1. COMPONENTES DE LA ONTOLOGÍA. Aunque esta es la presentación habitual, no es obligatoria. Por ejemplo, no es seguida en las clases de GEOP.

A efectos aclaratorios, cuando el vocabulario reutiliza elementos de otras herramientas, la descripción de estos es precedida por “Heredada de ____”. Asimismo, si se mencionan varios elementos, irán precedidos por el *qname* del vocabulario origen. La Tabla 1 muestra las propiedades *dayOfWeek* y *factor*, de la ontología EVENT. La primera, es heredada de TIME y la segunda específica de EVENT.

Propiedad	Descripción
<i>dayOfWeek</i>	Heredada de TIME. Permite expresar la relación entre <i>time:DateTimeDescription</i> y un día concreto de la semana de entre las instancias de la clase <i>time:DayOfWeek</i>
<i>factor</i>	Relaciona <i>event:Event</i> y <i>event:Factor</i> (herramienta, instrumento, causa abstracta, etc.). Su equivalente, <i>event:hasFactor</i> , ha quedado obsoleta. Inversa: <i>event:factor_of</i>

Tabla 1: Ejemplos de descripción de elementos en los vocabularios analizados

SNAP sigue una estructura algo diferente, al ser varias micro-ontologías las que lo conforman. También PROLES, ontología más compleja, se organiza distinguiendo elementos propios y heredados de otros vocabularios. En el caso de schema.org y DUL se han descrito los elementos que se han considerado más relevantes.

SCHEMA.ORG

Schema.org es una colección de esquemas para marcar páginas HTML que usan motores de búsqueda como Bing, Google Yahoo! y Yandex para mejorar la visualización y recuperación de recursos. Su utilidad reside precisamente en la extensión de HTML en todo el mundo, lo que favorece el uso de schema.org en toda página web, aplicando cualquiera de los tres formatos disponibles: Microdata, RDFa o JSON-LD. Este último aporta un formato más simple para estructurar grafos y que se mantiene separado del contenido de la página (Stuart, 2016, pp. 44-46). Existen multitud de esquemas y aunque el más general, *Thing*, puede aplicarse a cualquier tipo de recurso.

Schema.org es actualizado permanentemente. La última actualización en RDF analizada es realizada por Holger Knublauch el 23/03/2017, consta de 589 clases, 860 propiedades y 114 individuos (*enumeration values*). Dado el tamaño de este vocabulario, se analizan los elementos relacionados con el propósito de esta memoria.

Por lo que respecta a las acciones, en un sentido amplio, existen dos elementos: *Event* y *Action*: *Event* está orientado a actuaciones, interpretaciones, algo programado. Para hechos noticiosos que tienen lugar en un *EspacioTemporal* y físico, programados o no, es útil el esquema *Action*. Una *Action* es realizada por un agente y unos participantes sobre un objeto. Opcionalmente una acción ocurre en una localización con ayuda de un instrumento inanimado. La ejecución de la acción produce un resultado.

Por el tamaño de schema.org, no se describen todos sus componentes, pero sí se destacan algunos de los más relevantes. La Tabla 2 recoge varias clases de schema.org:

Clase	Descripción
<i>Action</i>	Acción llevada a cabo por un agente directo y participantes indirectos sobre un objeto directo. Subclases: <i>AchieveAction</i> , <i>AssessAction</i> , <i>ConsumeAction</i> , <i>ControlAction</i> , <i>CreateAction</i> , <i>FindAction</i> , <i>InteractAction</i> , <i>MoveAction</i> , <i>OrganizeAction</i> , <i>PlayAction</i> , <i>SearchAction</i> , <i>TradeAction</i> , <i>TransferAction</i> , <i>UpdateAction</i>
<i>Agent</i>	Ser, animado o no, ejecuta directamente una acción. Sólo se relaciona con <i>Action</i> , aunque sus dos subclases, <i>Person</i> y <i>Organization</i> , sí se utilizan con otros elementos
<i>Event</i>	Ocurre en un tiempo y lugar concreto, como un concierto, lectura o festival. Los Eventos repetidos deben ser estructurados como objetos <i>Event</i> separados. Subclases: <i>BusinessEvent</i> , <i>ChildrenEvent</i> , <i>SocialEvent</i> y <i>UserInteraction</i>
<i>Intangible</i>	Paraguas para cosas intangibles, como cantidades, valores estructurales, etc.
<i>Organization</i>	Organización, como una escuela, una ONG, corporación, club, etc. Algunas subclases de interés son: <i>GovernmentOrganization</i> , <i>LocalBusiness</i> , <i>Corporation</i>
<i>Person</i>	Una persona (viva, muerta, no muerta o ficticia)
<i>Place</i>	Entidades con extensión física y fija. Subclases: <i>AdministrativeArea</i> , <i>CivicStructure</i> , <i>Landform</i> , <i>LandmarksOrHistoricalBuildings</i> , <i>LocalBusiness</i> , <i>Residence</i> , <i>TouristAttraction</i>
<i>Product</i>	Cualquier producto o servicio ofrecido

Tabla 2: Algunas clases de schema.org

Schema.org contempla numerosas relaciones, algunas de las cuáles son:

Propiedad	Descripción
<i>actionStatus</i>	Relaciona <i>Action</i> y <i>ActionStatusType</i> para indicar la disposición actual de la acción
<i>agentActionStatusType</i>	Relaciona <i>Action</i> y <i>Person</i> u <i>Organization</i> , ejecutante de la acción
<i>aggregateRating</i>	Relaciona <i>Event</i> y <i>AggregateRating</i> para indicar la valoración global de un ítem, basado en un conjunto de comentarios o calificaciones
<i>attendee</i>	Relaciona <i>Event</i> y una <i>Person</i> u <i>Organization</i> que asiste al evento
<i>doorTime</i>	Relaciona <i>Event</i> y <i>DateTime</i> para indicar la hora de comienzo de admisión

Propiedad	Descripción
<i>duration</i>	Relaciona <i>Event</i> y <i>Duration</i> para indicar la duración del ítem, en formato ISO 8601
<i>endDate</i>	Relaciona <i>Event</i> y <i>Date</i> para indicar fecha y hora de finalización de un ítem, en formato ISO 8601
<i>endTime</i>	Relaciona <i>Action</i> y <i>DateTime</i> para expresar la fecha fin de algo
<i>error</i>	Relaciona <i>Action</i> y <i>Thing</i> para acciones fallidas, información sobre la causa del error
<i>eventStatus</i>	Permite indicar el estado de un <i>Event</i> ; útil cuando es cancelado o reprogramado
<i>image</i>	Imagen de un ítem, puede ser una URL o un <i>ImageObject</i> . Por tanto, también se puede considerar como propiedad tipo dato
<i>inLanguage</i>	Relaciona <i>Event</i> y el idioma del contenido o de la actuación (<i>Text or Language</i>). Puede utilizarse el estándar IETF BCP 47
<i>instrument</i>	Relaciona <i>Action</i> y <i>Thing</i> para indicar el objeto con que un agente ejecuta la acción
<i>location</i>	Permite indicar la localización de un <i>Event</i> o una <i>Action</i> , relacionándolo con <i>Place or Postal Address</i> . Localización del evento, organización o acción
<i>mainEntityOfPage</i>	Relaciona <i>Thing</i> y URL o <i>Creative Work</i> para indicar la página u otro <i>Creative Work</i> para el que es la principal entidad a ser descrita. También es una propiedad tipo dato
<i>object</i>	Relaciona <i>Action</i> y un <i>Thing</i> que funciona como objeto (paciente) sobre el que es ejecutada la acción, cuyo estado permanece intacto o cambia
<i>offers</i>	Permite indicar la <i>Offer</i> para proporcionar el ítem
<i>organizer</i>	Relaciona <i>Event</i> y <i>Person</i> u <i>Organization</i> organizador del mismo
<i>participant</i>	Relaciona <i>Action</i> y <i>Person</i> u <i>Organization</i> , que participan en la acción indirectamente
<i>performer</i>	Relaciona <i>Event</i> y su ejecutante (<i>Person</i> u <i>Organization</i>)
<i>potentialAction</i>	Relaciona <i>Thing</i> y una <i>Action</i> potencial, que describe una acción idealizada en la que éste jugaría el rol de 'objeto'
<i>previousStartDate</i>	Relaciona <i>Event</i> y <i>Date</i> . Usada junto con <i>eventStatus</i> para reprogramar o cancelar eventos. Esta propiedad contiene la fecha de comienzo previamente programada. Se debe utilizar la propiedad <i>startDate</i> para indicar la nueva fecha de inicio del evento
<i>recordedIn</i>	Relaciona <i>Event</i> y el <i>CreativeWork</i> que captura todo o parte de este <i>Event</i> . Inversa: <i>recordedAt</i> .
<i>result</i>	Relaciona <i>Action</i> y un <i>Thing</i> que funciona como resultado de la acción
<i>subEvent</i>	Relaciona dos <i>Events</i> donde uno esParteDe otro <i>Event</i> . Inversa: <i>superEvent</i>
<i>superEvent</i>	<i>Event</i> del que es parte otro <i>Event</i> . Inversa: <i>SubEvent</i>
<i>target</i>	Relaciona <i>Action</i> y un <i>EntryPoint</i> , indicando un punto de entrada (<i>EntryPoint</i> , dentro de algún protocolo basado en web) objetivo para una acción
<i>typicalAgeRange</i>	Permite indicar el rango de edad típica esperada con la clase <i>Text</i>

Tabla 3: Parte de las relaciones de schema.org

Por lo que respecta a las propiedades tipo dato, se pueden destacar las siguientes:

Propiedad	Descripción
<i>additionalType</i>	Tipo adicional para un ítem, usado habitualmente para añadir más tipos específicos desde vocabularios externos en sintaxis de microdatos
<i>alternateName</i>	Un alias, en formato texto, para un ítem
<i>DateTime</i>	Combinación de fecha y hora. Contempla diversas instancias que son utilizadas con distintas clases en schema, tales como <i>startTime</i> o <i>endTime</i>
<i>description</i>	Breve descripción del ítem, tipo texto
<i>image</i>	Imagen de un ítem, puede ser una URL o un <i>ImageObject</i> . También se puede considerar como propiedad objeto
<i>mainEntityOfPage</i>	Relaciona <i>Thing</i> y una URL o <i>Creative Work</i> para indicar la página u otro <i>Creative Work</i> para el que esta cosa es la principal entidad a ser descrita. También es una propiedad tipo objeto
<i>name</i>	Permite indicar el nombre de un ítem
<i>review</i>	Permite indicar la reseña del ítem
<i>sameAs</i>	Lleva a la URL de una <i>Thing</i> que claramente indica la identidad de un ítem
<i>startDate</i>	Permite indicar fecha y hora de inicio del ítem, en formato ISO 8601 al relacionarla con <i>Date</i>
<i>startTime</i>	Similar a <i>startDate</i> , relaciona <i>Action</i> y una fecha de inicio de la <i>Action</i>
<i>typicalAgeRange</i>	Permite indicar el rango de edad típica esperada con la clase <i>Text</i>
<i>url</i>	Permite especificar la URL de un ítem

Tabla 4: Parte de las propiedades tipo dato de *schema.org*

Para identificar *schema.org* en un recurso en línea se puede utilizar *Structured Data Testing Tool* que, además, permite ver qué nivel del vocabulario se ha implementado. Otras herramientas útiles son *Schema Creator* y *Thing Generator* que, en sentido inverso, permiten crear el código en microdatos de algunos elementos (personas, productos, eventos, organizaciones, películas, libros, reseñas). La Tabla 61 muestra los enlaces de estos elementos.

Schema.org está siendo mapeado con otros vocabularios con los que tiene nexos en común. En la web previa de esta herramienta (hoy se apunta a otra), se recoge información sobre algunos de los mapeados en los que se está trabajando. Estos se muestran en la siguiente tabla.

Vocabulario	Dominio	Enlace
<i>BIBO</i>	Recursos bibliográficos	http://bibliontology.com/
<i>DBpedia</i>	Recursos en línea	http://www.w3.org/2003/01/geo/
<i>DC</i>	Recursos	http://dublincore.org/documents/dcmi-terms/
<i>FOAF</i>	Agentes	http://xmlns.com/foaf/spec/
<i>GoodRelations</i>	Ofertas de productos y productos	http://purl.org/goodrelations/

Vocabulario	Dominio	Enlace
SIOC	Comunidades en línea y posts	http://sioc-project.org/
WordNet	Léxico en inglés	http://wordnet.princeton.edu/

Tabla 5: Vocabularios que están siendo mapeados con *schema.org*

DOLCE+DnS ULTRALITE ONTOLOGY (DUL)

DUL (versión 3.31), fue desarrollada siguiendo la metodología OntoClean. Es una simplificación de *DOLCE Lite-Plus library* y *Descriptions and Situations* (DnS), que tiene como fin asegurar la interoperabilidad entre ontologías de nivel medio y bajo. DUL contiene 6 grandes clases (66 en total), 105 relaciones y 5 propiedades tipo dato.

Según Ghislain Ateazing, en comentario de revisor de LOV (10/04/2017), los autores indican que la última versión incorpora una restricción a *Quality* y otra a *Region*. El fin es asegurar que las restricciones de las cualidades de DOLCE siempre estén asociadas a un región y viceversa.

Por su tamaño, no puede ser descrita en su totalidad, aunque sí algunos de sus elementos, que se muestran en las Tablas 6, 7 y 8.

Clase	Descripción
<i>Abstract</i>	Entidad que no se puede ubicar en el espacio-tiempo. Subclases: <i>FormalEntity</i> , <i>Region</i>
<i>Action</i>	Evento con al menos un ejecutante, típicamente definida en un plan, flujo de trabajo, proyecto, etc. Superclase: <i>Event</i>
<i>Agent</i>	Objeto agentivo. Subclases: <i>Person</i> , <i>PhysicalAgent</i> , <i>SocialAgent</i>
<i>Amount</i>	Cantidad, independientemente de cómo se mida o compute. Superclase: <i>Abstract</i>
<i>Collection</i>	Contenedor para entidades que comparten algún aspecto en común. Subclases: <i>Collective</i> , <i>Configuration</i> , <i>TypeCollection</i>
<i>Collective</i>	Compuesto por <i>Agents</i> , pueden actuar como agentes aunque se asume que no lo son. Superclase: <i>Collection</i> . Disjunta: <i>SocialAgent</i>
<i>CollectiveAgent</i>	<i>SocialAgent actedBy Agents</i> , que son (y actúan como) miembros de un <i>Collective</i> , puede tener un rol que también lo es de estos agentes. Superclase: <i>Agent</i> . Subclases: <i>Community</i> y <i>Group</i> . Disjunta: <i>Collective</i>
<i>Community</i>	Comunidad. Superclase: <i>CollectiveAgent</i>
<i>Concept</i>	Objeto social, descrito mediante <i>Description</i> . Definido el concepto, puede ser utilizado en otra <i>Description</i> . Superclase: <i>SocialObject</i> . Subclases: <i>EventType</i> , <i>LocalConcept</i> , <i>Parameter</i> , <i>Role</i>
<i>Configuration</i>	<i>Collective</i> cuyos miembros forman una unidad, entidad compuesta, de acuerdo con cierto esquema, que puede representarse en una <i>Description</i> . Ej.: un libro

Clase	Descripción
<i>Description</i>	<i>Social object</i> que representa una conceptualización. Puede ser vista como un 'contexto descriptivo' que usa o define conceptos para crear la visión de un contexto relacional (<i>Situation</i>) fuera del conjunto de datos de observación. Superclase: <i>Social object</i> . Subclases: <i>Contract, Design, Diagnosis, Goal, Method, Narrative, Norm, Plan, Relation, Right, Theory</i>
<i>Design</i>	<i>Description</i> de una <i>Situation</i> , en términos de estructura y función, mantenida por una <i>Entity</i> por alguna razón. Superclase: <i>Description</i>
<i>DesignedArtifact</i>	<i>PhysicalArtifact</i> que es descrito mediante un <i>Design</i> . Superclase: <i>PhysicalArtifact</i>
<i>Entity</i>	Cualquier cosa, real, posible o imaginaria, sobre la que quien modela quiere hablar
<i>Event</i>	Evento, proceso o estado físico, social o mental, concebido de formas diferentes según su aspecto, agentividad o participantes. DUL no elige una perspectiva u otra, pues un evento, como situación observable, puede ser visto de distintas formas a la vez. Se indica que, la clasificación basada en participantes parece ser la más estable y apropiada para modelar problemas. Para cada perspectiva, DUL contempla varias clases. Así, para 1) los aspectos de un evento, explota la noción de situación (observable); 2) en el caso de la intencionalidad, el evento tampoco tiene identidades diferentes, aunque el análisis de la causalidad genere situaciones distintas con identidades diferentes; 3) en cuanto la perspectiva de participantes en el evento, la identidad de éste sí se ve afectada por qué participa. Subclases: <i>Action, Process</i>
<i>EventType</i>	Clasifica cómo debe ser interpretado, ejecutado, previsto, visto, etc. un <i>Event</i> . Superclase: <i>Concept</i>
<i>FormalEntity</i>	Para entidades definidas formalmente, independientemente del contexto social en que se usan, no localizadas en tiempo o espacio. También llamadas <i>entidades platónicas</i> . Incluye entidades lógicas y matemáticas. Superclase: <i>Abstract</i>
<i>Group</i>	Grupo. Superclase: <i>CollectiveAgent</i>
<i>InformationEntity</i>	Pieza de información. Subclases: <i>InformationObject, InformationRealization</i>
<i>InformationObject</i>	Pieza de información, independientemente de cómo se realice. Superclase: <i>InformationEntity, SocialObject</i>
<i>InformationRealization</i>	Realización concreta de un <i>InformationObject</i> , por ejemplo, el documento escrito que contiene el texto de una ley. Superclase: <i>InformationEntity</i>
<i>LocalConcept</i>	Concepto que es descrito exactamente por una <i>Description</i> . Superclase: <i>Concept</i>
<i>NaturalPerson</i>	Persona física, en la intuición del sentido común. Superclase: <i>PhysicalAgent, Person</i>
<i>Object</i>	Cualquier objeto físico, social o mental, o una sustancia. Algunas subclases: <i>Agent, PhysicalObject, SocialObject</i>
<i>Organism</i>	Objeto físico con características biológicas. Superclases: <i>PhysicalAgent, SocialAgent</i>
<i>Organization</i>	<i>SocialAgent</i> creado de forma convencional, estructurado internamente. Necesita de un <i>Role</i> y un <i>Agent</i> para actuar. Superclase: <i>SocialAgent</i>
<i>Parameter</i>	Concepto que clasifica una <i>Region</i> . Representa restricciones sobre valores observables (representados mediante <i>Region</i>). Superclase: <i>Concept</i>
<i>Person</i>	Persona. Superclase: <i>Agent</i> . Subclases: <i>NaturalPerson</i> y <i>SocialPerson</i>
<i>Personification</i>	Entidad con características agentivas, cuyo status es <i>ResultadoDe</i> una transformación cultural, como el Santo Grial. Superclase: <i>SocialAgent</i>

Clase	Descripción
<i>PhysicalAgent</i>	Objeto capaz de conceptualizar una <i>Description</i> para planificar una <i>Action</i> . Superclases: <i>Agent</i> , <i>PhysicalObject</i> . Subclases: <i>NaturalPerson</i> y <i>Organism</i>
<i>PhysicalArtifact</i>	Descrito mediante un <i>Plan</i> . Superclase: <i>PhysicalObject</i> . Subclases: <i>DesignedArtifact</i>
<i>PhysicalAttribute</i>	Valor físico para describir un objeto físico, como densidad, color, etc. Superclase: <i>Abstract</i>
<i>PhysicalBody</i>	Objetos físicos para los que tendemos a neutralizar cualquier característica artificial. Superclases: <i>PhysicalObject</i> . Subclases: <i>BiologicalObject</i> , <i>ChemicalObject</i> y <i>Substance</i>
<i>PhysicalObject</i>	Objeto que ocupa una región espacial. Subclases: <i>PhysicalAgent</i> , <i>PhysicalArtifact</i> , <i>PhysicalBody</i> , <i>PhysicalPlace</i>
<i>PhysicalPlace</i>	Objeto físico que está localizado intrínsecamente, por ejemplo, una zona de agua
<i>Place</i>	Localización, en un sentido muy general. Para una localización "absoluta" utilizar <i>SpaceRegion</i> . Superclase: <i>SocialObject</i>
<i>Process</i>	Evento en evolución o que, en todo caso, no dependen estrictamente de agentes, tareas y planes. Superclase: <i>Event</i>
<i>Quality</i>	Cualidad de un objeto, no una parte de éste, sin la cual no existe
<i>Region</i>	Espacio dimensional que puede ser utilizado como valor para una cualidad de una entidad. Superclase: <i>Abstract</i> . Subclases: <i>Amount</i> , <i>PhysicalAttribute</i> , <i>SocialAttribute</i> , <i>SpaceRegion</i> , <i>SpatioTemporalRegion</i> , <i>TimeInterval</i>
<i>Role</i>	Concepto que clasifica un <i>Object</i> . Superclase: <i>Concept</i>
<i>Situation</i>	Una vista, consistente con ('satisfacer') la <i>Description</i> de un conjunto de entidades. También puede ser vista como un 'contexto relacional' creado por un observador sobre la base de un 'marco' (esto es, una <i>Description</i>). Requiere satisfacer alguna <i>Description</i> . Superclase: <i>SocialObject</i> .
<i>SocialAgent</i>	Individuo cuya existencia deriva de su comunicabilidad social y capacidad de acción. Requiere de un <i>PhysicalAgent</i> para actuar. Superclase: <i>SocialObject</i> . Subclases: <i>CollectiveAgent</i> , <i>Organization</i> , <i>Personification</i> , <i>SocialPerson</i>
<i>SocialAttribute</i>	Región en un espacio dimensional que representa alguna característica de un <i>SocialObject</i> como valores de juicio o escalas sociales, etc. Superclase: <i>Region</i>
<i>SocialObject</i>	Cualquier objeto que sólo existe en la comunicación de un <i>Event</i> en el que al menos un <i>PhysicalObject</i> participa. Necesita ser expresado mediante algún <i>InformationObject</i> . Subclases: <i>Collection</i> , <i>Concept</i> , <i>Description</i> , <i>InformationObject</i> , <i>Place</i> , <i>Situation</i> , <i>SocialAgent</i>
<i>SocialPerson</i>	<i>SocialAgent</i> que necesita la existencia de una <i>NaturalPerson</i> específica para actuar. Superclase: <i>SocialAgent</i> y <i>Person</i>
<i>SpaceRegion</i>	Región en un espacio dimensional que permite localizar una entidad. A diferencia de <i>Place</i> , tiene un espacio dimensional específica. Superclase: <i>Region</i>
<i>SpatioTemporalRegion</i>	Región constituida por un <i>SpaceRegion</i> and a <i>TimeInterval</i> . Superclase: <i>Region</i>
<i>TimeInterval</i>	Región en un espacio dimensional y representa el tiempo. Superclase: <i>Region</i>
<i>TypeCollection</i>	Colección cuyos miembros son el conjunto máximo de personas que comparten el mismo tipo, por ejemplo, "los italianos". Superclase: <i>Collection</i>

Tabla 6: Algunas clases de la ontología DUL

DUL incluye varias propiedades tipo dato, entre las que destacan las de la Tabla 7.

Propiedad	Descripción
<i>associatedWith</i>	Propiedad principal, de la que penden las demás
<i>actsFor</i>	Vincula <i>Agent</i> y un <i>Social agent</i> . Tiene como propiedad inversa <i>actsThrough</i>
<i>classifies</i>	Clasifica cualquier <i>Entity</i> como <i>Concept</i>
<i>co-participatesWith</i>	Relaciona dos <i>Objects</i> que participan en un mismo <i>Event</i>
<i>conceptualizes</i>	Relación un <i>Agent</i> y la representación interna de un <i>SocialObject</i> . Por ejemplo, 'John cree en la teoría de la conspiración'. Las descripciones y conceptos pueden ser 'asumidas', las situaciones pueden ser 'cree' o 'conocidas', los planes pueden ser "adoptados", etc. Tiene como relación inversa <i>isConceptualizedBy</i>
<i>express</i>	Relaciona un <i>InformationObject</i> y un <i>SocialObject</i>
<i>follows</i>	Relaciona secuencialmente dos entidades. Su relación inversa es <i>precedes</i>
<i>hasConstituent</i>	Relaciona dos entidades tal que la segunda es <i>ParteDe</i> la primera, en un sentido de pertenencia no partonómico. Por ejemplo, podemos hablar de elementos físicos que forman parte de elementos no físicos, como las personas que constituyen un sistema. Para relaciones taxonómicas, existe la relación <i>hasPart</i> . Inversa: <i>isPartOf</i>
<i>hasLocation</i>	Relación genérica entre <i>Entities</i> para ubicar una en otra. Para localizaciones "absolutas", hay que utilizar <i>SpaceRegion</i> . Inversa: <i>isLocationOf</i>
<i>hasParticipant</i>	Relaciona un <i>Event</i> y los objetos que participan en el mismo. Inversa: <i>isParticipantIn</i>
<i>hasQuality</i>	Expresa la relación entre una <i>Entity</i> y una <i>Quality</i> . Inversa: <i>isQualityOf</i>
<i>hasRegion</i>	Expresa la relación entre una <i>Entity</i> y una <i>Region</i> . Inversa: <i>isRegionFor</i>
<i>hasSetting</i>	Expresa la relación entre una <i>Entity</i> y una <i>Situation</i> . Inversa: <i>isSettingFor</i> . Subpropiedades: <i>hasEventIncludedIn</i> , <i>hasObjectIncludedIn</i> , <i>hasTimeIncludedIn</i>
<i>introducing</i>	Relaciona <i>Description</i> y <i>SocialAgent</i> . Por ejemplo, una Carta Constitucional <i>introduces</i> Presidente de la República
<i>overlaps</i>	Expresa la relación entre dos <i>Entities</i> que se superponen
<i>realizes</i>	Relaciona <i>InformationRealization</i> e <i>InformationObject</i> . Por ejemplo, la copia en papel de la Constitución italiana "hace posible" el texto de la Constitución. Inversa: <i>isRealizedBy</i> . Relacionada con ésta <i>realizesInformationAbout</i> , vincula <i>InformationRealization</i> con una <i>Entity</i> . Por ejemplo, relaciona la copia en papel de la Constitución italiana e Italia. Inversa: <i>isReferenceOfInformationRealizedBy</i>
<i>satisfies</i>	Expresa la relación entre una <i>Situation</i> y una <i>Description</i> . Inversa: <i>isSatisfiedBy</i>
<i>specializes</i>	Funciona de forma similar a las relaciones entre clases y subclases, entre <i>SocialObjects</i> . Por ejemplo, el rol <i>PhDStudent</i> <i>specializes</i> el rol <i>Student</i> . Inversa: <i>isSpecializedBy</i>
<i>usesConcept</i>	Relación genérica entre <i>Description</i> y un <i>Concept</i> . Para ser utilizado, un <i>Concept</i> debe ser previamente <i>definedIn</i> (subpropiedad de <i>usesConcept</i>) otra <i>Description</i> . Inversa: <i>isConceptUsedIn</i>

Tabla 7: Algunas relaciones de la ontología DUL

También se definen estas propiedades tipo dato:

Propiedad	Descripción
<i>hasDataValue</i>	<p>Propiedad principal, de la que penden el resto. Formas de codificar los valores:</p> <ol style="list-style-type: none"> 1. Simplemente añadir un valor utilizando <i>hasDataValue</i> 2. Definir una <i>Region</i> para una <i>Entity</i> usando <i>hasRegion</i>, y después añadir un valor para esa <i>Region</i> a través de <i>hasRegionDataValue</i> 3. Definir una <i>Quality</i> para una <i>Entity</i> mediante <i>hasQuality</i>; vincular dicha <i>Quality</i> con una <i>Region</i> y añadir un valor a esa <i>Region</i> mediante la propiedad <i>hasRegionDataValue</i> 4. Cuando un valor es obligatorio pero no se ha observado directamente, se puede asignar un valor a un <i>Parameter</i> con <i>hasParameterDataValue</i> y asociar dicho <i>Parameter</i> a una entidad utilizando <i>is constraint for</i> 5. Cuando un valor es obligatorio pero no se ha observado directamente, también se puede definir un <i>Parameter</i> para una <i>Region</i> y asignar un valor a esa <i>Region</i>, mediante el uso de <i>hasRegionDataValue</i> <p>Los cinco enfoques obedecen a diferentes requisitos. El primero es el más simple, pero si hace falta establecer un intervalo entre dos valores se debe utilizar el segundo. Además, si hay que distinguir una cualidad individual de un valor, se debe recurrir al patrón 3. Los patrones 4 y 5 se pueden utilizar cuando se modela una restricción o selección independientemente de la observación real de valores en el mundo real.</p>
<i>hasEventDate</i>	Asigna un valor tipo fecha a un <i>Event</i> . Es repetible, pues puede asignarse fechas diferentes: fecha de inicio, fecha fin, intervalo, etc. Además, el formato puede variar: fecha completa, año, fecha y hora, etc.
<i>hasParameterDataValue</i>	Parametriza valores para un <i>Parameter</i>
<i>hasRegionDataValue</i>	Codificar valores para una <i>Region</i> . Subpropiedad: <i>hasIntervalDate</i> , que permite asignar valores, en diferentes formatos, a la clase <i>TimeInterval</i> . Es repetible

Tabla 8: Propiedades tipo dato en DUL

En esta ontología se reutilizan las siguientes herramientas:

Vocabulario	Dominio	Enlace
<i>DOLCE Lite-Plus Library</i>	Fundacional	http://www.w3.org/2001/sw/BestPractices/WNET/DLP3941_daml.html
<i>Patrón DNS</i>	Descripciones y situaciones	http://www.loa.istc.cnr.it/ontologies/ExtendedDnS.owl

Tabla 9: Herramientas reutilizadas en DUL

THE EVENT ONTOLOGY (EVENT)

Desarrollada por el *Centre for Digital Music*, de la Universidad de Londres, para la descripción de eventos. La concepción de un evento depende de cómo sea percibido por un agente cognitivo, una persona, que clasifica una región espacio/temporal como tal (Raimond y Abdallah, 2007). El mundo no contiene eventos, si no que son patrones de cambio identificados por los agentes, en los que pueden participar activamente agentes, factores pasivos y productos que se localizan en un espacio físico y uno temporal.

EVENT se compone de las siguientes clases:

Clase	Descripción
<i>Agent</i>	Heredada de FOAF. Agente activo (una persona, un ordenador, ...)
<i>DateTimeDescription</i>	Heredada de TIME. Describe la fecha u hora de un <i>DateTimeInterval</i> mediante la relación <i>hasDateTimeDescription</i> . Subclase: <i>time:January</i>
<i>DateTimeInterval</i>	Heredada de TIME. Define un intervalo de tiempo. Única <i>TemporalEntity</i> que puede ser descrita a partir de <i>DateTimeDescription</i> . Superclase: <i>ProperInterval</i>
<i>DayOfWeek</i>	Heredada de TIME. Permite indicar un día de la semana a partir de siete instancias, utilizando la relación <i>dayOfWeek</i>
<i>DurationDescription</i>	Heredada de TIME. Duración de una <i>TemporalEntity</i> . Subclase: <i>time:Year</i>
<i>Event</i>	Clasificación arbitraria de una región espacio/temporal por un agente cognitivo
<i>Factor</i>	Herramienta, instrumento, causa abstracta, etc., que puede usarse en un evento. Se constituye a partir de la propiedad <i>factor_of</i> y algún elemento de la clase <i>Event</i> . Es una clase definida, que no forma parte del modelo como tal, no se utiliza directamente (RaimondAbdallah, 2007)
<i>Instant</i>	Heredada de TIME. Entidad temporal sin extensión. Superclase: <i>TemporalEntity</i>
<i>Interval</i>	Heredada de TIME. Entidad temporal con extensión. Superclase: <i>TemporalEntity</i> . Subpropiedad: <i>ProperInterval</i>
<i>January</i>	Heredada de TIME. Enero. Superclase: <i>time:DateTimeDescription</i>
<i>Person</i>	Heredada de FOAF. Sin relaciones con otros elementos <i>foaf:Agent</i>
<i>Product</i>	Cualquier cosa producida por un evento, constituida a partir de la propiedad <i>produces_in</i> y algún elemento de la clase <i>Event</i> . Es una clase definida, que no forma parte del modelo como tal, no se utiliza directamente (Raimond y Abdallah, 2007)
<i>SpatialThing</i>	Heredada de GEO. Algo con extensión espacial (tamaño, forma o posición), por ejemplo, personas, lugares o áreas abstractas como cubos
<i>TemporalEntity</i>	Heredada de TIME. [Momento en que se produce un evento]. Subclases: <i>time:Instant</i> y <i>time:Interval</i>
<i>TimeZone</i>	Heredada de TZONT para indicar la zona horaria de una <i>TemporalEntity</i>
<i>Year</i>	Heredada de TIME. Superclase: <i>time:DurationDescription</i>

Tabla 10: Clases de la ontología EVENT

EVENT se compone de las siguientes relaciones o propiedades tipo objeto:

Propiedad	Descripción
<i>after</i>	Heredada de TIME. Relaciona dos <i>time:TemporalEntity</i> tal que el dominio de la propiedad ocurre después que su rango. Propiedad inversa: <i>time:before</i>
<i>agent</i>	Relaciona <i>event:Event</i> y <i>foaf:Agent</i> . Su equivalente <i>event:hasAgent</i> ha quedado obsoleta (<i>deprecated</i>). Inversa: <i>event:agent_in</i>
<i>agent_in</i>	Relaciona un <i>foaf:Agent</i> y un <i>event:Event</i> . Su equivalente <i>event:isAgentIn</i> ha quedado obsoleta (<i>deprecated</i>). Inversa: <i>event:agent</i>
<i>before</i>	Heredada de TIME. Relaciona dos <i>time:TemporalEntity</i> , el dominio de la propiedad ocurre antes que su rango. Propiedad inversa: <i>time:after</i>
<i>dayOfWeek</i>	Heredada de TIME. Expresa la relación entre <i>time:DateTimeDescription</i> y un día concreto de la semana, instancias de <i>time:DayOfWeek</i>
<i>factor</i>	Relaciona <i>event:Event</i> y <i>event:Factor</i> (herramienta, instrumento, causa abstracta,

Propiedad	Descripción
	etc.). Su equivalente, <i>event:hasFactor</i> , es obsoleta. Inversa: <i>event:factor_of</i>
<i>factor_of</i>	Relaciona <i>event:Factor</i> y <i>event:Event</i> . Su equivalente, <i>event:isFactorOf</i> , ha quedado obsoleta. Inversa: <i>event:factor</i>
<i>hasBeginning</i>	Heredada de TIME. Indica el comienzo de <i>time:TemporalEntity</i> usando un <i>time:Instant</i> como objeto
<i>hasDateTimeDescription</i>	Heredada de TIME. Relaciona <i>time:DateTimeInterval</i> y <i>time:DateTimeDescription</i>
<i>hasDurationDescription</i>	Heredada de TIME. Relaciona <i>time:TemporalEntity</i> y <i>time:DurationDescription</i>
<i>hasEnd</i>	Heredada de TIME. Indica el final de una <i>time:TemporalEntity</i> usando <i>time:Instant</i> como objeto
<i>inDateTime</i>	Heredada de TIME. Indica el momento en que se produce un <i>time:Instant</i> , relacionando esta clase con <i>time:DateTimeDescription</i>
<i>inside</i>	Heredada de TIME. Relaciona <i>time:Instant</i> e <i>time:Interval</i> , situando al 1º en el 2º
<i>intervalAfter</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> , tal que el dominio de la propiedad ocurre después que su rango. Inversa: <i>time:intervalBefore</i>
<i>intervalBefore</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> , su dominio ocurre antes que su rango. Superpropiedad: <i>time:Before</i> . Inversa: <i>time:intervalAfter</i>
<i>intervalContains</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> . Inversa: <i>time:intervalDuring</i>
<i>intervalDuring</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> . Inversa: <i>time:intervalContains</i>
<i>intervalEquals</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i>
<i>intervalFinishedBy</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> indicando que una (dominio) es acabada por otra (rango). Inversa: <i>time:intervalFinishes</i>
<i>intervalFinishes</i>	De TIME, relaciona dos <i>ProperInterval</i> . Inversa: <i>time:intervalFinishedBy</i>
<i>intervalMeets</i>	De TIME, relaciona dos <i>time:ProperInterval</i> . Inversa: <i>time:intervalMetBy</i>
<i>intervalMetBy</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> . Inversa: <i>time:intervalMeets</i>
<i>intervalOverlappedBy</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> . El rango de la propiedad se superpone a su dominio. Propiedad inversa: <i>time:intervalOverlaps</i>
<i>intervalOverlaps</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> . El dominio de la propiedad se superpone a su rango. Inversa: <i>time:intervalOverlappedBy</i>
<i>intervalStartedBy</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> . El rango de la propiedad da inicio a su dominio. Inversa: <i>time:intervalStarts</i>
<i>intervalStarts</i>	Heredada de TIME. Relaciona dos <i>time:ProperInterval</i> . El dominio de la propiedad da inicio a su rango. Propiedad inversa: <i>time:intervalStartsBy</i>
<i>maker</i>	Heredada de FOAF. Relaciona una cosa realizada (<i>Thing</i>) por <i>foaf:Agent</i> y éste
<i>place</i>	Relaciona <i>event:Event</i> y <i>geo:SpatialThing</i>
<i>product</i>	Relaciona <i>event:Event</i> y su resultado. Su equivalente, <i>event:hasProduct</i> , ha quedado obsoleta. Inversa: <i>event:produce_in</i>
<i>produced_in</i>	Relaciona el resultado de la ejecución de un evento y dicho <i>event:Event</i> . Su equivalente, <i>event:producedIn</i> , ha quedado obsoleta. Inversa: <i>event:produce_in</i>

Propiedad	Descripción
<i>sub_event</i>	Propiedad reflexiva para <i>event:Event</i> . Aporta una forma de dividir un evento en partes más simples. Su equivalente, <i>event:hasSubEvent</i> , ha quedado obsoleta
<i>time</i>	Relaciona <i>event:Event</i> y <i>time:TemporalEntity</i> , clasificando una región temporal
<i>timeZone</i>	Heredada de TIME. Permite expresar la relación entre <i>time:DateTimeDescription</i> y una zona horaria (<i>timezone:TimeZone</i>)
<i>unitType</i>	Heredada de TIME. Relaciona <i>time:DateTimeDescription</i> y cualquier unidad temporal contenida en <i>time:TemporalUnit</i>

Tabla 11: Propiedades tipo objeto de la ontología EVENT

Algunas propiedades de EVENT han quedado obsoletas, por ejemplo *hasAgent*, *isAgentIn*. Se han definido propiedades equivalentes, que se indican en la Tabla 11.

EVENT se compone de las siguientes propiedades tipo dato:

Propiedad	Descripción
<i>date</i>	Heredada de DCMES. Punto o periodo de tiempo asociado a un evento en el ciclo de vida de un recurso
<i>day</i>	Permite indicar el día concreto en que tiene lugar un <i>time:DateTimeDescription</i>
<i>dayOfYear</i>	Permite indicar el día concreto del año en que tiene lugar un <i>time:DateTimeDescription</i>
<i>days</i>	Indica la duración de <i>time:TemporalEntity</i> a través de <i>time:DurationDescription</i>
<i>description</i>	Heredada de DCMES. Descripción de un recurso
<i>hour</i>	Permite indicar la hora concreta en que tiene lugar un <i>time:DateTimeDescription</i>
<i>hours</i>	Indica la duración de <i>time:TemporalEntity</i> a través de la clase <i>time:DurationDescription</i>
<i>inXSDDateTime</i>	Permite expresar, de forma más simple que <i>time:DateTimeDescription</i> , datos de tiempo para <i>time:Instant</i>
<i>literal_factor</i>	Relaciona <i>event:Event</i> y un factor, descrito como literal. Su equivalente, <i>event:hasLiteralFactor</i> ha quedado obsoleta
<i>minute</i>	Permite indicar el minuto concreto en que tiene lugar un <i>time:DateTimeDescription</i>
<i>minutes</i>	Indica la duración de <i>time:TemporalEntity</i> a través de la clase <i>time:DurationDescription</i>
<i>month</i>	Permite indicar el mes concreto en que tiene lugar un <i>time:DateTimeDescription</i>
<i>months</i>	Indica la duración de <i>time:TemporalEntity</i> a través de la clase <i>time:DurationDescription</i>
<i>second</i>	Permite indicar el segundo concreto en que tiene lugar un <i>time:DateTimeDescription</i>
<i>seconds</i>	Indica la duración de <i>time:TemporalEntity</i> a través de <i>time:DurationDescription</i>
<i>title</i>	Heredada de DCMES. Nombre dado a un recurso
<i>week</i>	Permite indicar la semana concreta en que tiene lugar un <i>time:DateTimeDescription</i>
<i>weeks</i>	Indica la duración de <i>time:TemporalEntity</i> a través de <i>time:DurationDescription</i>
<i>xsdDateTime</i>	Expresa, más simple que <i>time:DateTimeDescription</i> , datos de tiempo para <i>time:Interval</i>
<i>year</i>	Permite indicar el año concreto en que tiene lugar un <i>time:DateTimeDescription</i>
<i>years</i>	Indica la duración de <i>time:TemporalEntity</i> a través de <i>time:DurationDescription</i>

Tabla 12: Propiedades tipo dato de la ontología EVENT

Gráficamente, EVENT es representada mediante la siguiente figura⁵⁵.

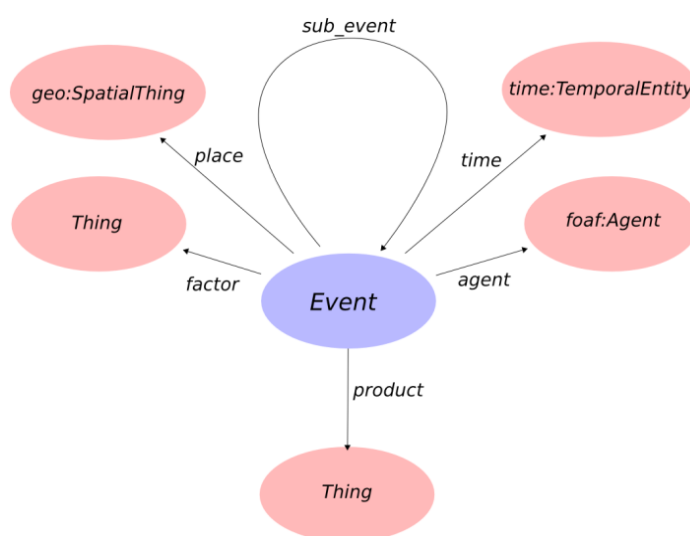


Figura 77: Modelo de Event

En la Figura 77 puede observarse la mayor parte de los vocabularios reutilizados por Event, detallados también en la Tabla 13.

Vocabulario	Dominio	Enlace
DCMES	Recursos	http://dublincore.org/documents/dces/
GEO	Espacios físicos	http://www.w3.org/2003/01/geo/
FOAF	Agentes	http://xmlns.com/foaf/spec/
TIME	Espacios temporales	http://www.w3.org/TR/owl-time/
TZONT	Zonas horarias	http://www.w3.org/2006/timezone

Tabla 13: Vocabularios reutilizados en EVENT

FRIEND OF A FRIEND (FOAF)

Friend of a Friend (FOAF), versión 0.99, es una especificación RDF/OWL para identificar agentes (personas, organizaciones, grupos), definir algunas de sus características y relaciones, con el fin de vincular agentes e información en línea sobre estos. Por vínculos se entiende:

Redes sociales de colaboración humana, amistad y asociación; Redes representacionales que describen una visión simplificada de un universo en dibujos en términos factuales; Redes de información que usan vínculos basados en web para compartir descripciones publicadas independientemente de este mundo interconectado (Brickley y Miller, 2014).

⁵⁵ Fuente: <http://motools.sourceforge.net/event/event.html> (último acceso: 20/05/2017)

Actualmente está formada por 13 clases y 62 propiedades, con distintos grados de estabilidad, que se indica mediante un *status* para cada término: *stable* (6 clases, 10 propiedades), *unstable* (4 clases, 5 propiedades), *testing* (3 clases, 39 propiedades), *archaic* (8 propiedades). FOAF trabaja a dos niveles, *core* (identificado en Tabla 14, Tabla 15 y Tabla 16 con (C)), con los elementos básicos de descripción de un agente, independientemente de tiempo y tecnología; y *web social* (W), con elementos característicos de los agentes en redes sociales. La especificación también contempla otros elementos, que se identifican con (O). A continuación se describen las clases de FOAF.

Clase	Descripción
<i>Agent</i> (C)	Agente (persona, un grupo, un software, etc.), cosas (<i>things</i>) que hacen cosas (<i>Stuff</i>). Asimilada con <i>dcterms:Agent</i> . Subclases: <i>Group</i> , <i>Organization</i> , <i>Person</i>
<i>Document</i> (C)	Cosas que, en sentido general, se entienden como documentos. Asimilada con <i>schema:CreativeWork</i> . Subclase: <i>Image</i>
<i>Group</i> (C)	Colección de agentes individuales. Los miembros que componen esta clase también son <i>Agent</i> , vinculados con <i>Group</i> mediante la relación <i>member</i>
<i>Image</i> (C)	Documentos que son imágenes. Superclase: <i>Document</i>
<i>LabelProperty</i> (W)	Se define como propiedad para RDF con valor de texto que sirve como etiqueta
<i>OnlineAccount</i> (W)	Indica la cuenta de un <i>Agent</i> en algún tipo de servicio en línea. Subclases: <i>OnlineEcommerceAccount</i> , <i>OnlineGameAccount</i> , <i>OnlineChatAccount</i>
<i>OnlineChatAccount</i> (W)	Cuenta en servicio de mensajería en línea de <i>Agent</i> . Superclase: <i>OnlineAccount</i>
<i>OnlineEcommerceAccount</i> (W)	Cuenta en servicio de comercio online de un <i>Agent</i> . Superclase: <i>OnlineAccount</i>
<i>OnlineGameAccount</i> (W)	Cuenta en servicio de juego en línea de un <i>Agent</i> . Superclase: <i>OnlineAccount</i>
<i>Organization</i> (C)	Tipo de agente para instituciones sociales, como empresas, sociedades, etc. Superclase: <i>Agent</i>
<i>Person</i> (C)	Una persona. Asimilada con la clase <i>Person</i> de SCHEMA.ORG y SWAP (<i>Semantic web Application Platform</i>)
<i>PersonalProfileDocument</i> (W)	Documentos que usan RDF para describir propiedades de la persona que ha elaborado dicho documento. Las propiedades <i>maker</i> y <i>primaryTopic</i> de este documento tienen como objeto la misma <i>Person</i>
<i>Project</i> (C)	Proyecto, en sentido amplio, colectivo o individual, formal o informal. Puede que en un futuro se relacione con las propiedades <i>currentProject</i> y <i>pastProject</i>
<i>SpatialThing</i> (W)	Heredada de GEO para indicar algo con extensión espacial, que es cualquier lugar. En esta ontología, es la superclase de <i>Person</i>

Tabla 14: Clases de FOAF

Se muestran las relaciones de FOAF. No incluye los elementos arcaicos (*dnaChecksum*, *family_name*, *fundedBy*, *geekcode*, *givenname*, *holdsAccount*, *surname*, *theme*).

Propiedad	Descripción
<i>account</i> (W)	Relaciona un <i>Agent</i> y una <i>OnlineAccount</i>
<i>accountServiceHomepage</i> (W)	Relaciona <i>OnlineAccount</i> y la página web del suministrador del servicio, que se indica a partir de la clase <i>Document</i>

Propiedad	Descripción
<i>basedNear</i> (C)	Localización de algo cercano, de acuerdo con la noción humana de cercanía. Relaciona dos <i>geo:SpatialThing</i>
<i>currentProject</i> (W)	Relaciona <i>Person</i> y un proyecto en el que trabaja actualmente, cualquier <i>Thing</i>
<i>depiction</i> (W)	Relaciona una cosa (<i>Thing</i>) y su representación mediante <i>Image</i> . Suele utilizarse para indicar el contenido de una imagen. Subpropiedad: <i>img</i> . Inversa: <i>depicts</i>
<i>depicts</i> (W)	Relaciona <i>Image</i> y la cosa que representa (<i>Thing</i>). Suele utilizarse para indicar el contenido de una imagen. Inversa: <i>depicts</i>
<i>focus</i> (W)	Heredada de SKOS. Relaciona el <i>skos:Concept</i> de algo y ese algo (<i>Thing</i>)
<i>homepage</i> (W)	Relaciona una cosa (<i>Thing</i>) y una página de inicio sobre ésta (<i>Document</i>). Superpropiedades: <i>page</i> y <i>isPrimaryTopicOf</i>
<i>img</i> o <i>image</i> (W)	Relaciona una cosa (<i>Thing</i>) y una representación particular de ésta mediante la clase <i>Image</i> , por ejemplo, una foto de perfil. Superpropiedad: <i>depiction</i>
<i>interest</i> (W)	Apunta al <i>Document</i> sobre un tema/s de interés para un <i>Agent</i>
<i>isPrimaryTopicOf</i> (C)	Relaciona <i>Thing</i> y <i>Document</i> , que tiene como tema dicha cosa. Superpropiedad: <i>page</i> . Inversa: <i>primaryTopic</i>
<i>jabberID</i> (W)	Relaciona <i>Agent</i> y un identificador textual asignado al mismo en Jabber
<i>knows</i> (C)	Relaciona <i>Person</i> conocida por otra <i>Person</i> . Implica cierto grado de reciprocidad
<i>logo</i> (W)	Logo (<i>Thing</i>) que representa alguna cosa (<i>Thing</i>)
<i>made</i> (W)	Relaciona <i>Agent</i> y lo que realiza (<i>Thing</i>). Inversa: <i>maker</i>
<i>maker</i> (C)	Relaciona una cosa realizada (<i>Thing</i>) por <i>Agent</i> y éste. Inversa: <i>made</i>
<i>mbox</i> (W)	Relaciona <i>Agent</i> y una cuenta de correo personal en Internet de la que es propietario (<i>Thing</i>), identificada habitualmente mediante The mailto URL scheme ⁵⁶
<i>member</i> (C)	Permite relacionar un <i>Group</i> y sus miembros (<i>Agent</i>)
<i>openid</i> (W)	Propiedad funcional que relaciona <i>Agent</i> y un <i>Document</i> , que es usado como identificador indirecto de <i>Agent</i> , de forma similar al “Identity URL” de OpenID. Superpropiedad: <i>isPrimaryTopicOf</i>
<i>page</i> (W)	Relaciona <i>Thing</i> y <i>Document</i> (página web) sobre ésta. Subpropiedades: <i>homepage</i> , <i>isPrimaryTopicOf</i> , <i>tipjar</i> , <i>weblog</i> . Inversa: <i>topic</i>
<i>pastProject</i> (W)	Relaciona <i>Person</i> y proyecto en el que trabajó anteriormente, cualquier <i>Thing</i>
<i>phone</i> (O)	Permite indicar un teléfono utilizando Addressing Schemes, hoy retirado
<i>primaryTopic</i> (C)	Propiedad funcional que relaciona un <i>Document</i> y la cosa (<i>Thing</i>) principal sobre la que trata. Inversa: <i>isPrimaryTopicOf</i>
<i>publications</i> (W)	Relaciona <i>Person</i> y sus publicaciones (<i>Document</i>), habitualmente publicadas a través de su página de inicio
<i>schoolHomepage</i> (W)	Relaciona <i>Person</i> y <i>Document</i> , que es la página de inicio de su escuela
<i>thumbnail</i> (W)	Relaciona una <i>Image</i> (a tamaño completo) y su miniatura (clase <i>Image</i>)
<i>tipjar</i> (W)	Relaciona <i>Agent</i> y <i>Document</i> que describe cómo compensar a ese agente por algo, por ejemplo, en el intercambio de contenidos en línea. Superpropiedad: <i>page</i>
<i>topic</i> (W)	Relaciona <i>Document</i> y la cosa (<i>Thing</i>) sobre la que trata. Inversa: <i>page</i>

56 <https://www.ietf.org/rfc/rfc2368.txt> (último acceso: 20/05/2017)

Propiedad	Descripción
<i>topic_interest (W)</i>	Relaciona <i>Agent</i> y un tema que es de interés para éste (<i>Thing</i>). A diferencia de <i>topic</i> , relaciona directamente con el interés, y no con el <i>Document</i> sobre éste
<i>weblog (W)</i>	Relaciona <i>Agent</i> y un <i>weblog</i> de dicho agente
<i>workInfoHomepage (W)</i>	Relaciona <i>Person</i> y su lugar de trabajo, indicado a través de un <i>Document</i>
<i>workplaceHomepage (W)</i>	Relaciona <i>Person</i> y la web de su trabajo, indicado a través de un <i>Document</i>

Tabla 15: Relaciones en FOAF

Estas son las propiedades tipo dato de FOAF:

Propiedad	Descripción
<i>account (W)</i>	Permite relacionar un <i>Agent</i> y una <i>OnlineAccount</i>
<i>accountName (W)</i>	Indica, mediante un literal, el nombre (identificador) asociado con una <i>OnlineAccount</i>
<i>age (W)</i>	Permite indicar la fecha de cumpleaños de un <i>Agent</i> , con formato mm-dd
<i>aimChatID (O)</i>	Indica un identificador textual para <i>Agent</i> en el sistema de mensajería AOL (AIM). Superpropiedad: <i>nick</i> . Propiedad objeto y dato
<i>birthday (O)</i>	Expresa el día de cumpleaños de <i>Agent</i> con una cadena string
<i>familyName (W)</i>	Permite indicar el apellido familiar de una <i>Person</i>
<i>firstName (W)</i>	Permite indicar el primer nombre de una <i>Person</i>
<i>gender (W)</i>	Indica género de <i>Agent</i> , habitualmente femenino o masculino, pero no necesariamente
<i>givenName (W)</i>	Permite indicar el apellido familiar de una <i>Person</i>
<i>icqChatID (O)</i>	Indica un identificador textual para <i>Agent</i> en el sistema de mensajería instantánea ICQ. Superpropiedad: <i>nick</i> . En el archivo de la ontología también aparece como relación
<i>jabberId (W)</i>	Indicar mediante un literal, el jabberID de algo
<i>lastName (W)</i>	Permite indicar el segundo nombre de una <i>Person</i>
<i>mbox_sha1sum (W)</i>	Indica, para <i>Agent</i> el sha1sum de un nombre URI de una cuenta de correo personal. En la ontología aparece también como una relación
<i>MSNChatID (O)</i>	Indica un identificador textual para <i>Agent</i> en MSN. Superpropiedad: <i>nick</i> . Es propiedad objeto y dato
<i>myersBriggs (W)</i>	Relaciona <i>Person</i> y un literal que consiste en un valor extraído de la clasificación de la personalidad Myers Briggs (MBTI)
<i>name (C)</i>	Permite indicar el nombre de un <i>Thing</i> en cadena de texto
<i>nick (W)</i>	Nickname breve e informal con el que se caracteriza un <i>Agent</i> . Subpropiedades: <i>AIMChatID</i> , <i>ICQChatID</i> , <i>MSNChatID</i> , <i>SkypeID</i> , <i>YahooChatID</i>
<i>plan (W)</i>	Permite comentar el plan de <i>Person</i> en el protocolo finger ⁵⁷ y el archivo .plan
<i>sha1 (W)</i>	Indica, para un <i>Document</i> la forma textual de un “SHA1 hash” (función de cifrado) de alguna representación de su contenido
<i>skypeChatID (O)</i>	Indica un identificador textual para <i>Agent</i> en Skype. Superpropiedad: <i>nick</i>
<i>status (O)</i>	Expresa, mediante una cadena de texto, algo por lo que un <i>Agent</i> es feliz para el público en general, en relación con su actividad actual, humor, localización, etc.

57 http://www.rajivshah.com/Case_Studies/Finger/Finger.htm (último acceso: 20/05/2017)

Propiedad	Descripción
<i>title (C)</i>	Título de una persona (Mr, Mrs, Ms, Dr.). Está prevista su clasificación como obsoleta
<i>yahooChatID (O)</i>	Indica identificador textual para <i>Agent</i> en Yahoo. Superpropiedad: <i>nick</i> . Es propiedad objeto y dato

Tabla 16: Propiedades tipo dato de FOAF

FOAF nace en el 2000, como formato no propietario, disponible en RDF. Aunque no es un estándar W3C, aprovecha gran parte de los esfuerzos de normalización e interoperabilidad de esta entidad. Es muy utilizado, especialmente entre los comunidad *Open Source*. Con la aplicación de Javascript *FOAF-a-matic*, introduciendo en un formulario unos datos básicos, se genera automáticamente código RDF. FOAF reutilizan vocabularios como GEO para *SpatialThing*, y establece equivalencias con Dublin Core (*Agent*, propiedad *Maker*) y schema.org (*Document*, *Image*, *Person*) (ver Tabla 17).

Vocabulario	Dominio	Enlace
GEO	Espacios físicos	http://www.w3.org/2003/01/geo/
DCTERMS	Recursos	http://dublincore.org/documents/dcmi-terms/
SCHEMA.ORG	Recursos en Internet	http://schema.org/

Tabla 17: Vocabularios reutilizados en FOAF

TIME

Ontología del *Information Science Institute* (ISI), de la University of Southern California para definir conceptos temporales. En febrero de 2017 se publicó el borrador de una nueva versión de esta ontología⁵⁸. No se ha tenido en cuenta en esta descripción.

Merece la pena destacar dos clases: *Interval* e *Instant*, entidades temporales con y sin extensión temporal, respectivamente. Ambas se relacionan mediante *inside* (dominio: *Interval*; rango: *Instant*). Sus elementos se resumen en las siguientes tablas.

Clases de la ontología TIME:

Clase	Descripción
<i>DateTimeDescription</i>	Describe una fecha u hora de un <i>DateTimeInterval</i> . Para ello utiliza la relación <i>hasDateTimeDescription</i>
<i>DateTimeInterval</i>	Define un intervalo de tiempo. Única <i>TemporalEntity</i> que puede ser descrita a partir de <i>DateTimeDescription</i> . Superclase: <i>ProperInterval</i> Define un intervalo mediante <i>hasDateTimeDescription</i> y <i>DateTimeDescription</i> y recurre a propiedades tipo dato (<i>day</i> , <i>hour</i> , <i>minute</i> , <i>month</i> , <i>second</i> , <i>week</i> , <i>year</i>). Estos pueden ser de dos tipos: a) <i>xsd:gday</i> , <i>xsd:gmonth</i> , <i>xsd:gyear</i> (la “g” hace referencia al calendario gregoriano); b) valores enteros no negativos (<i>nonNegativeInteger</i>) para semanas, horas, minutos y segundos
<i>DayOfWeek</i>	Indica un día de la semana a partir de siete instancias usando la relación <i>dayOfWeek</i>

58 Borrador TIME Ontology: <https://www.w3.org/TR/owl-time/> (último acceso: 01/06/2017)

Clase	Descripción
<i>DurationDescription</i>	Duración de una <i>TemporalEntity</i> Definida a través de una serie de restricciones descriptivas, permite múltiples descripciones de la duración (máximo 1 literal para cada una de las propiedades tipo dato <i>days</i> , <i>hours</i> , <i>minutes</i> , <i>months</i> , <i>seconds</i> , <i>weeks</i> , <i>years</i>), expresadas en base decimal (<i>xsd:decimal</i>)
<i>Instant</i>	Entidad temporal sin extensión. Superclase: <i>TemporalEntity</i> Para indicar el momento en que se produce un <i>Instant</i> se utiliza <i>inDateTime</i> y como objeto <i>DateTimeDescription</i> ó, más simple, expresar una fecha u hora mediante propiedades dato con <i>xsdDateTime</i> (para intervalos) y <i>inXSDDateTime</i> (para instantes)
<i>Interval</i>	Entidad temporal con extensión. Superclase: <i>TemporalEntity</i> . Subpropiedad: <i>ProperInterval</i>
<i>ProperInterval</i>	Intervalos cuya fecha de inicio y fin son diferentes. Superpropiedad: <i>Interval</i> . Subpropiedad: <i>DateTimeInterval</i>
<i>TemporalEntity</i>	Entidad temporal. Subclases: <i>Instant</i> e <i>Interval</i> La duración de una entidad temporal (<i>TemporalEntity</i>) puede definirse mediante un cómputo general o indicando dos fechas entre las que tiene lugar
<i>TemporalUnit</i>	Clase para definir unidades de tiempo, desde <i>unitDay</i> a <i>unitYear</i>
<i>TimeZone</i>	Heredada de TZONT para indicar la zona horaria en que tiene lugar <i>TemporalEntity</i>

Tabla 18: Clases de la ontología TIME

Las relaciones de la ontología TIME se indican en la tabla que sigue:

Propiedad	Descripción
<i>after</i>	Relaciona dos <i>TemporalEntity</i> , tal que el dominio de la propiedad ocurre después que su rango. Propiedad inversa: <i>before</i>
<i>before</i>	Relaciona dos <i>TemporalEntity</i> , tal que el dominio de la propiedad ocurre antes que su rango. Propiedad inversa: <i>after</i>
<i>dayOfWeek</i>	Permite expresar la relación entre <i>DateTimeDescription</i> y un día concreto de la semana de entre las instancias de la clase <i>DayOfWeek</i>
<i>hasBeginning</i>	Indica el comienzo de una <i>TemporalEntity</i> utilizando un <i>Instant</i> como objeto
<i>hasDateTimeDescription</i>	Relaciona una <i>DateTimeInterval</i> con un <i>DateTimeDescription</i>
<i>hasDurationDescription</i>	Relaciona una <i>TemporalEntity</i> con una <i>DurationDescription</i>
<i>hasEnd</i>	Indica el final de una <i>TemporalEntity</i> utilizando un <i>Instant</i> como objeto
<i>inDateTime</i>	Indica cuándo se produce <i>Instant</i> , relacionándolo con <i>DateTimeDescription</i>
<i>inside</i>	Relaciona <i>Instant</i> e <i>Interval</i> , situando al primero en el segundo
<i>intervalAfter</i>	Relaciona dos <i>ProperInterval</i> , de forma que el dominio de la propiedad ocurre después que su rango. Propiedad inversa: <i>intervalBefore</i>
<i>intervalBefore</i>	Relaciona dos <i>ProperInterval</i> , de forma que el dominio de la propiedad ocurre antes que su rango. Propiedad inversa: <i>intervalAfter</i>
<i>intervalContains</i>	Relaciona dos <i>ProperInterval</i> . Inversa: <i>intervalDuring</i>
<i>intervalDuring</i>	Relaciona dos <i>ProperInterval</i> . Inversa: <i>intervalContains</i>
<i>intervalEquals</i>	Relaciona dos <i>ProperInterval</i>

Propiedad	Descripción
<i>intervalFinishedBy</i>	Relaciona dos <i>ProperInterval</i> indicando que una (dominio) es acabada por otra (rango). Inversa: <i>intervalFinishes</i>
<i>intervalFinishes</i>	Relaciona dos <i>ProperInterval</i> . Inversa: <i>intervalFinishedBy</i>
<i>intervalMeets</i>	Relaciona dos <i>ProperInterval</i> . Inversa: <i>intervalMetBy</i>
<i>intervalMetBy</i>	Relaciona dos <i>ProperInterval</i> . Inversa: <i>intervalMeets</i>
<i>intervalOverlappedBy</i>	Relaciona dos <i>ProperInterval</i> . El rango de la propiedad se superpone a su dominio. Propiedad inversa: <i>intervalOverlaps</i>
<i>intervalOverlaps</i>	Relaciona dos <i>ProperInterval</i> . El dominio de la propiedad se superpone a su rango. Inversa: <i>intervalOverlappedBy</i>
<i>intervalStartedBy</i>	Relaciona dos <i>ProperInterval</i> . El rango de la propiedad da inicio a su dominio. Inversa: <i>intervalStarts</i>
<i>intervalStarts</i>	Relaciona dos <i>ProperInterval</i> . El dominio de la propiedad da inicio a su rango. Propiedad inversa: <i>intervalStartsBy</i>
<i>timeZone</i>	Heredada de TIMEZONE. Relaciona <i>DateTimeDescription</i> y <i>timezone:TimeZone</i>
<i>unitType</i>	Relaciona <i>DateTimeDescription</i> y la unidad temporal contenida en <i>TemporalUnit</i>

Tabla 19: Relaciones de la ontología TIME

Finalmente, estas son las propiedades tipo dato de TIME:

Propiedad	Descripción
<i>day</i>	Indica el día concreto en que tiene lugar un <i>DateTimeDescription</i>
<i>dayOfYear</i>	Indica el día concreto del año en que tiene lugar un <i>DateTimeDescription</i>
<i>days</i>	Indica la duración de una <i>TemporalEntity</i> a través de la clase <i>DurationDescription</i>
<i>hour</i>	Indica la hora concreta en que tiene lugar un <i>DateTimeDescription</i>
<i>hours</i>	Indica la duración de una <i>TemporalEntity</i> a través de la clase <i>DurationDescription</i>
<i>inXSDDateTime</i>	Expresa, de forma más simple que <i>DateTimeDescription</i> , datos de tiempo para <i>Instant</i>
<i>minute</i>	Indica el minuto concreto en que tiene lugar un <i>DateTimeDescription</i>
<i>minutes</i>	Indica la duración de una <i>TemporalEntity</i> a través de la clase <i>DurationDescription</i>
<i>month</i>	Indica el mes concreto en que tiene lugar un <i>DateTimeDescription</i>
<i>months</i>	Indica la duración de una <i>TemporalEntity</i> a través de la clase <i>DurationDescription</i>
<i>second</i>	Indica el segundo concreto en que tiene lugar un <i>DateTimeDescription</i>
<i>seconds</i>	Indica la duración de una <i>TemporalEntity</i> a través de la clase <i>DurationDescription</i>
<i>week</i>	Indica la semana concreta en que tiene lugar un <i>DateTimeDescription</i>
<i>weeks</i>	Indica la duración de una <i>TemporalEntity</i> a través de la clase <i>DurationDescription</i>
<i>xsdDateTime</i>	Expresa, de forma más simple que <i>DateTimeDescription</i> , datos de tiempo para <i>Interval</i>
<i>year</i>	Permite indicar el año concreto en que tiene lugar un <i>DateTimeDescription</i>
<i>years</i>	Indica la duración de una <i>TemporalEntity</i> a través de la clase <i>DurationDescription</i>

Tabla 20: Propiedades tipo dato de la ontología TIME

Es importante tener en cuenta las diferencias horarias entre distintas zonas del mundo ya que, si la ontología es utilizada en diferentes países, es muy probable que las fechas y horas no coincidan. (Hobbs y Pan, 2006) resaltan esta cuestión en la especificación, distinguiendo “zonas horarias” de “estándares de tiempo” (*time standards*), que son utilizados en esas zonas horarias. Un estándar de tiempo es una forma de calcular el tiempo con respecto a un sistema mundial de cálculo. Para esta cuestión, se creó TZONT, que combina elementos geográficos y temporales para definir una hora.

Vocabulario	Dominio	Enlace
TZONT	Zonas horarias	http://www.w3.org/2006/timezone

Tabla 21: Vocabularios reutilizados en TIME

TIME se puede descargar en la web del *Information Science Institute* (ISI), de la University of Southern California, en formato OWL. Recientemente, en febrero de 2017, se ha publicado un borrador de la ontología, que prevé la inclusión de nuevos términos, como *GeneralDateTimeDescription*.

WGS84 GEO POSITIONING (GEO)

En un eje de coordenadas imaginario (ver Figura 78⁵⁹), la latitud, dada en grados relativos al Ecuador, se establece con el eje de las Y, mientras que la longitud, dada en grados del meridiano de Greenwich, se establece con el eje de las X.

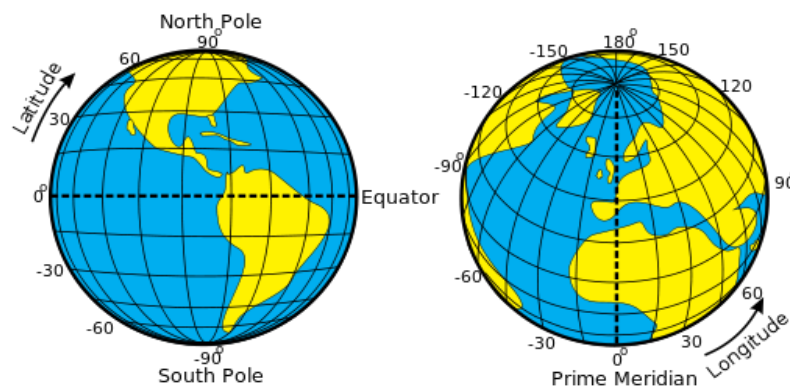


Figura 78: Latitud y longitud del planeta Tierra

GEO permite ubicar un punto concreto en un espacio de la superficie terrestre, teniendo en cuenta estas coordenadas a partir de dos clases, una propiedad tipo objeto y cuatro de anotación. Por lo que respecta sus clases, tiene dos, recogidas en la siguiente tabla:

⁵⁹ Latitud y longitud de la Tierra. Wikimedia:
https://commons.wikimedia.org/wiki/File:Latitude_and_Longitude_of_the_Earth_mk.svg (último acceso: 01/06/2017)

Clase	Descripción
<i>Point</i>	Define un punto a partir de la asignación de coordenadas que definen latitud, longitud y altura de un punto, habitualmente mediante la especificación WGS84 (<i>World Geodetic System 1984</i>). Superclase: <i>SpatialThing</i>
<i>SpatialThing</i>	Cualquier cosa con extensión espacial (tamaño, forma o posición). Subclase: <i>Point</i>

Tabla 22: Clases de GEO

Relaciones de GEO:

Propiedad	Descripción
location	Relaciona algo y el punto (<i>SpatialThing</i>) u otra cosa geométrica en el espacio donde se encuentra. Por ejemplo, la relación entre una torre de radio y un punto con determinada lat y longitud, o la relación entre un parque y su contorno como un arco cerrado de puntos. Superclase: <i>foaf:based_near</i>

Tabla 23: Relaciones de GEO

En GEO no se definen propiedades tipo dato, pero sí de anotación:

Propiedad	Descripción
alt	Permite indicar la altitud de un <i>SpatialThing</i> en formato WGS84 (en grados decimales)
lat	Permite indicar la latitud de un <i>SpatialThing</i> en formato WGS84 (en grados decimales)
lat_long	Una representación, separada por comas, de una coordenada de latitud y longitud
long	Permite indicar la longitud de un <i>SpatialThing</i> en formato WGS84 (en grados decimales)

Tabla 24: Propiedades de anotación de GEO

Esta ontología no reutiliza otros vocabularios, aunque sí indica que la propiedad *location* es una subclase de *foaf:based_near*. Fue revisada por última vez en 2009, de acuerdo con la descripción del vocabulario⁶⁰.

FAO GEOPOLITICAL ONTOLOGY (GEOPOLITICAL)

Food and Agriculture Organization of the United Nations (FAO) diseñó GEOPOLITICAL para facilitar el intercambio y compartición de datos sobre países y/o regiones de forma estandarizada. La ontología maneja información en varios idiomas (español, inglés, francés, árabe, ruso, chino mandarín e italiano), mapea diferentes sistemas de codificación estándar (UN, ISO, FAOSTAT, AGROVOC, etc.), indica relaciones entre territorios (fronteras, pertenencia a grupos, etc.) y refleja sus cambios históricos. Ha sido poblada con datos de fuentes reconocidas como la propia FAO, Naciones Unidas (UN), International Organization for Standardization (ISO) o Banco Mundial (Kim, Iglesias-Sucasas, y Viollier, 2013; Kim et al., 2013).

⁶⁰ Vocabulario GEO: http://www.w3.org/2003/01/geo/wgs84_pos# (último acceso: 20/05/2017)

Aplicada a la instancia *Spain*, pertenece a la clase *self_governing* y *hasBorderWith* con *Andorra*, *France*, *Gibraltar*, *Morocco*, *Portugal*. Además, *isInGroup* por cuestiones económicas (*EU*), geográficas (*Europe*, *southern Europe*), en organizaciones (*FAO*) y en un grupo especial (*World*). La Figura 79⁶¹ muestra las relaciones de la instancia *Russian Federation*.

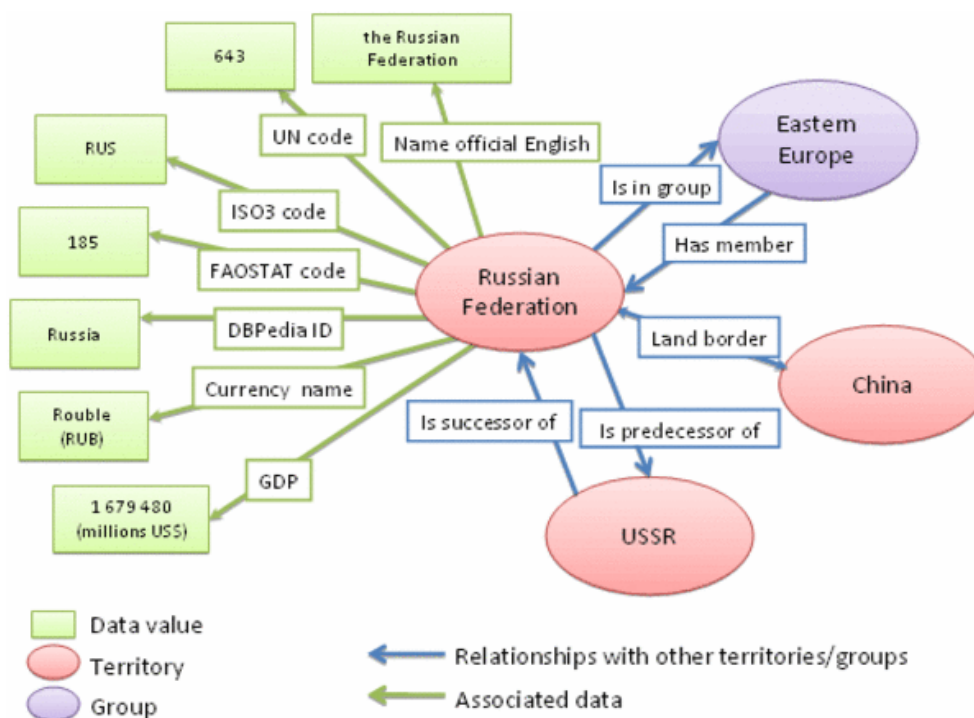


Figura 79: Clases, relaciones y propiedades para "Russian Federation" en GEOP

GEOPOLITICAL contiene once clases, seis relaciones, más de ochenta propiedades de tipo dato y más de trescientas instancias. Las clases reflejan distintas divisiones del mundo, físicas, políticas y económicas, entre otras. Las siguientes tablas se basan en la información del fichero owl del vocabulario.

Clase	Descripción
<i>area</i>	Área. Subclases: <i>territory</i> y <i>group</i>
<i>territory</i>	Superficie de la Tierra. Superclase: <i>area</i> . Subclases: <i>disputed</i> , <i>self_governing</i> , <i>non_self_governing</i> , <i>other</i>
<i>disputed</i>	Territorios en disputa. Superclase: <i>territory</i>
<i>self_governing</i>	Países con gobierno propio, según UNML ⁶² . Superclase: <i>territory</i>

61 Vocabulario GEOPOLITICAL: <http://www.fao.org/countryprofiles/geoinfo/en/> (último acceso: 20/05/2017)

62 United Nation Map Library (UNLM), Department of Public Information (DPI), Dag Hammarskjöld Library (DHL) de las Naciones Unidas: <http://research.un.org/en/maps> (último acceso: 20/05/2017)

Clase	Descripción
<i>non_self_governing</i>	Países que carecen de gobierno propio, de acuerdo con la sección de Cartografía, Department of Field Support de Naciones Unidas ⁶³ . Superclase: <i>territory</i>
<i>other</i>	Otro tipo de territorio, no clasificado antes. Superclase: <i>territory</i>
<i>group</i>	Algún tipo de agrupación de territorios. Superclase: <i>area</i> . Subclases: <i>organization</i> , <i>geographical_region</i> , <i>special_group</i> , <i>economic_region</i>
<i>organization</i>	Organización. Superclase: <i>group</i>
<i>geographical_region</i>	Regiones geográficas en que se divide el mundo, según <i>United Nation Statistics Division</i> ⁶⁴ . Superclase: <i>group</i>
<i>special_group</i>	Grupos especiales. Superclase: <i>group</i>
<i>economic_region</i>	Regiones económicas en que se divide el mundo. Superclase: <i>group</i>

Tabla 25: Clases de GEOP

Relaciones de la ontología GEOP:

Propiedad	Descripción
<i>hasBorderWith</i>	Relaciona 2 <i>territory</i> que son frontera entre sí, de acuerdo con la sección de Cartografía, Department of Field Support de Naciones Unidas
<i>hasMember</i>	Establece una relación entre <i>group</i> y <i>territory</i> . Inversa: <i>isInGroup</i>
<i>isAdministeredBy</i>	Relaciona un <i>non_self_governing</i> con otro que lo administra (<i>self_governing</i>)
<i>isInGroup</i>	Establece una relación entre <i>territory</i> y <i>group</i> . Inversa: <i>hasMember</i>
<i>IsPredecessorOf</i>	Relaciona 2 áreas, permitiendo realizar un seguimiento de los cambios históricos en territorios. Inversa: <i>isSuccessorOf</i>
<i>isSuccessorOf</i>	Relaciona 2 áreas, indicando que el dominio de esta relación es, históricamente hablando, posterior a su rango. Inversa: <i>IsPredecessorOf</i>

Tabla 26: Propiedades tipo objeto de GEOP

Algunas de las propiedades tipo dato de GEOPOLITICAL son:

Propiedad	Descripción
<i>hasCode</i>	Clasifica un área mediante algún código internacional. Tiene varias subpropiedades, de acuerdo a la codificación internacional: <i>codeAGROVOC</i> , <i>codeDBPedia</i> , <i>codeUN</i> , <i>codeFAOSTAT</i> , <i>codeFAOTERM</i> , <i>codeGAUL</i> , <i>codeISO2</i> , <i>codeISO3</i> , <i>codeUNDP</i>
<i>hasCoordinate</i>	Ubica geográficamente un territorio de acuerdo con sus coordenadas geográficas (latitud y longitud máximas y mínimas). Subpropiedades: <i>hasMaxLatitude</i> , <i>hasMaxLongitude</i> , <i>hasMinLatitude</i> , <i>hasMinLongitude</i>

63 Sección de Cartografía, Department of Field Support de Naciones Unidas:

<http://www.un.org/Depts/Cartographic/map/profile/world00.pdf> (último acceso: 20/05/2017)

64 United Nation Statistics Division: <http://unstats.un.org/unsd/methods/m49/m49regin.htm> (último acceso: 20/05/2017)

Propiedad	Descripción
<i>hasCurrency</i>	Indica la/s moneda/s circulante/s de un territorio. Para ello, se vale de 8 subpropiedades: <i>codeCurrency</i> , que asigna un código a la moneda, de acuerdo con FAO TERM PORTAL ⁶⁵ . El resto de términos muestra el nombre de la moneda en diferentes idiomas: <i>nameCurrencyAR</i> , <i>nameCurrencyEN</i> , <i>nameCurrencyES</i> , <i>nameCurrencyFR</i> , <i>nameCurrencyIT</i> , <i>nameCurrencyRU</i> , <i>nameCurrencyZH</i>
<i>hasListName</i>	Define un nombre para el área, en los diferentes idiomas con que trabaja la ontología: <i>nameListNameAR</i> , etc.
<i>hasNationality</i>	Define la nacionalidad de un área, de acuerdo con FAO TERM PORTAL, en los diferentes idiomas con que trabaja la ontología: <i>nationalityAR</i> , etc.
<i>hasOfficialName</i>	Permite indicar el nombre oficial de un área, de acuerdo con FAO TERM PORTAL, en los diferentes idiomas con que trabaja la ontología: <i>nameOfficialAR</i> , etc.
<i>hasShortName</i>	Permite indicar el nombre corto de un área, de acuerdo con FAO TERM PORTAL, en los diferentes idiomas con que trabaja la ontología: <i>nameShortAR</i> , etc.
<i>hasStatistics</i>	Agrupar varias subpropiedades para reflejar datos estadísticos e indicadores sobre un área. Comprende 6 propiedades: <i>agriculturalArea</i> , <i>countryArea</i> , <i>GDP</i> , <i>HDI (Human Development Index)</i> , <i>landArea</i> , <i>population</i> . Estas propiedades se dividen en otras 4 que permiten indicar el año del que se obtiene el dato, cifra total, unidad de medida y notas. Sus datos proceden de la base de datos de FAOSTAT ⁶⁶
<i>validSince</i>	Indica el primer año de validez de un área particular, desde 1985. Si tiene como valor 1985 significa que es válida desde esa fecha o antes
<i>validUntil</i>	Indica el último año de validez de un área particular. Si tiene como valor 9999 significa que aún es válida

Tabla 27: Algunas propiedades tipo dato de la ontología GEOP

Esta ontología no reutiliza otros vocabularios. Mucha de esta información está disponible en la web de la FAO para cada perfil de país (*FAO Country Profile*)⁶⁷.

BBCCORE (CORE CONCEPTS ONTOLOGY)

BBCCORE es una ontología genérica, de la BBC, para definir cosas tratadas habitualmente por la empresa: personas, lugares, eventos, organizaciones y temas. El objetivo es facilitar la descripción de piezas informativas y vincular los conceptos de clientes de la BBC con los definidos en la ontología a través de la propiedad *rdfs:subClassOf*.

Este vocabulario forma parte de plataforma Linked Data de la BBC y está interrelacionada con otras ontologías de la BBC. Conocer los objetivos de éstas facilita entender cómo funciona BBCCORE:

1. *BBC Ontology (BBC)*: codifica la lógica que conecta documentos web, productos de la BBC y plataformas para las que el contenido está disponible;
2. *Creative Work Ontology (CWORK)* indica el número mínimo de metadatos

65 FAO TERM PORTAL: <http://www.fao.org/faoterm/en/> (último acceso: 20/05/2017)

66 FAOSTAT, alojados en <http://faostat.fao.org/DesktopDefault.aspx?PageID=377&lang=en#anchor> hasta el 23/10/2014 y en <http://faostat3.fao.org/home/E> posteriormente (último acceso: 20/05/2017)

67 FAO Country Profile: <http://www.fao.org/countryprofiles/en/> (último acceso: 20/05/2017)

necesarios para expresar una pieza de contenido creativo en la BBC;

3. *CMS ontology (CMS)* permite describir el sistema de gestión de contenidos de la empresa;
4. *Provenance Ontology (PROVENANCE)*: permite definir tipos de gráficos almacenados por la BBC y asociarlos a metadatos, lo que facilita la gestión, validación y exposición de datos en los servicios de la empresa.

La Figura 80⁶⁸ muestra varios elementos de ontologías de la BBC.

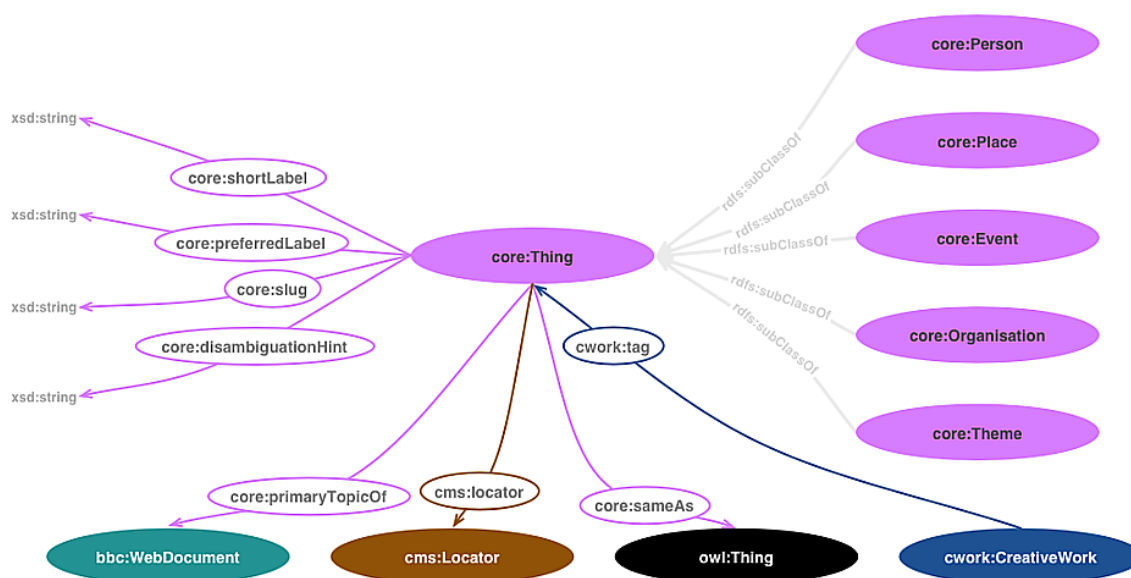


Figura 80: Diagrama de la ontología de la BBC

BBCCORE está compuesta por 33 clases, 36 relaciones y 21 propiedades tipo dato, entre otros elementos. La Tabla 28 recoge sus clases, incluyendo aquellas que reutiliza, principalmente de DCTERMS, por ejemplo:

Clase	Descripción
<i>Agent</i>	Heredada de DCTERMS. Recurso que actúa o tiene poder para actuar.
<i>AgentClass</i>	Heredada de DCTERMS. Un grupo de agentes
<i>BibliographicResource</i>	Heredada de DCTERMS. Un libro, artículo u otro recurso documental.
<i>Event</i>	Clase genérica para eventos (musicales, históricos, deportivos, políticos, etc.)
<i>FileFormat</i>	Heredada de DCTERMS. Formato de un recurso digital. Superclase: <i>dcterms:MediaType</i>
<i>Frequency</i>	Heredada de DCTERMS. Ritmo al que algo se repite

68 BBC CORE CONCEPTS: <http://www.bbc.co.uk/ontologies/coreconcepts> (último acceso: 20/05/2017)

Clase	Descripción
<i>Jurisdiction</i>	Heredada de DCTERMS. Extensión o rango de una autoridad judicial, fuerzas de la ley u otra. Superclase: <i>dcterms:LocationPeriodOrJurisdiction</i>
<i>LicenseDocument</i>	Heredada de DCTERMS. Documento legal que otorga un permiso oficial para hacer algo con el recurso. Superclase: <i>dcterms:RightsStatement</i>
<i>LinguisticSystem</i>	Heredada de DCTERMS. Sistema de signos, símbolos, sonidos, gestos o reglas utilizados en comunicación.
<i>Location</i>	Heredada de DCTERMS. Región espacial o lugar con nombre: <i>dcterms:LocationPeriodOrJurisdiction</i>
<i>LocationPeriodOrJurisdiction</i>	Heredada de DCTERMS. Localización, un periodo de tiempo o jurisdicción. Subclases: <i>Jurisdiction, Location, PeriodOfTime</i>
<i>Locator</i>	Heredada de CMS. Representa los identificadores requeridos para localizar la representación de algo en otro sistema de la BBC. Aparece en el diagrama de la ontología
<i>MediaType</i>	Heredada de DCTERMS. Formato de archivo o medio físico. Superclase: <i>dcterms:MediaTypeOrExtent</i> . Subclases: <i>dcterms:FileFormat y dcterms:PhysicalMedium</i>
<i>MediaTypeOrExtent</i>	Heredada de DCTERMS. Tipo de medio o extensión. Subclases: <i>dcterms:MediaType y dcterms:SizeOrDuration</i>
<i>MethodOfAccrual</i>	Heredada de DCTERMS. Método de incorporación de recursos a la colección
<i>MethodOfInstruction</i>	Heredada de DCTERMS. Proceso utilizado para generar conocimiento, actitudes y habilidades
<i>Ontology</i>	Heredada de PROVENANCE
<i>Organisation</i>	Clase para definir organización en la BBC. Por ejemplo, empresas
<i>PeriodOfTime</i>	Heredada de DCTERMS. Intervalo de tiempo que es nombrado o definido por sus fechas de inicio y fin. Superclase: <i>dcterms:LocationPeriodOrJurisdiction</i>
<i>Person</i>	Clase para personas en la BBC. Subclases: políticos, atletas, figuras históricas, contribuidores en programas, etc.
<i>PhysicalMedium</i>	Heredada de DCTERMS. Material físico. Superclase: <i>dcterms:MediaType</i>
<i>PhysicalResource</i>	Heredada de DCTERMS. Una cosa material
<i>Place</i>	Clase para lugares en la BBC
<i>Policy</i>	Heredada de DCTERMS. Plan o curso de acción por una autoridad, con el fin de influir y determinar decisiones, acciones y otros asuntos
<i>ProvenanceState</i>	Heredada de DCTERMS. Declaración de cualquier cambio en la propiedad y custodia de un recurso desde su creación y que resulta significativo para su autenticidad, integridad e interpretación
<i>RightsStatement</i>	Heredada de DCTERMS. Declaración de los derechos de propiedad intelectual de un recurso, documento legal que otorga un permiso oficial para hacer algo con el recurso, o una declaración sobre derechos de acceso. Subclase: <i>dcterms:LicenseDocument</i>
<i>SizeOrDuration</i>	Heredada de DCTERMS. Una dimensión o extensión, o el tiempo tomado para ejecutar algo. Superclase: <i>dcterms:MediaTypeOrExtent</i>

Clase	Descripción
<i>Standard</i>	Heredada de DCTERMS. Base para una comparación; punto de referencia contra el que otras cosas pueden ser evaluadas
<i>Theme</i>	Concepto utilizado para cosas intangibles o conceptos abstractos como crisis financiera obstetricia, felicidad
<i>Thing</i>	Superclase de todas las cosas que la BBC crea o sobre la que habla
<i>VocabularyEncodingScheme</i>	Heredada de DCAM, indica la procedencia de un vocabulario. Incluye qué individuos (vocabularios) admite: DCMIType, DDC, IMT, LCC, LCSH, MESH, NLM, TGN, UDC
<i>webDocument</i>	Heredada de BBC Ontology, indica la web a la que se apunta

Tabla 28: Clases de la ontología BBCCORE

Relaciones de BBCCORE:

Propiedad	Descripción
<i>accessRights</i>	Heredada de DCTERMS. Permite indicar quién puede acceder a determinado recurso o una indicación sobre su estado de seguridad. Rango: <i>dcterms:RightsStatement</i>
<i>accrualMethod</i>	De DCTERMS. Indica <i>dcterms:MethodOfAccrual</i> de adición de un ítem a la colección
<i>accrualPeriodicity</i>	De DCTERMS. Indica <i>dcterms:Frequency</i> de adición de un ítem a la colección
<i>accrualPolicy</i>	De DCTERMS. Indica <i>dcterms:Policy</i> por la que se añade un ítem a la colección
<i>audience</i>	Heredada de DCTERMS. Indica la <i>dcterms:AgentClass</i> para quién está destinado un recurso o resulta útil
<i>conformsTo</i>	Heredada de DCTERMS. Indica el <i>dcterms:Standard</i> establecido para que se ajuste el recurso descrito
<i>contributor</i>	De DCTERMS. Indica <i>dcterms:Agent</i> responsable de hacer contribuciones al recurso
<i>coverage</i>	Heredada de DCTERMS. Indica el tema espacial o temporal de un recurso, su aplicabilidad espacial o la jurisdicción bajo la cuál dicho recurso es relevante, teniendo como rango <i>dcterms:LocationPeriodOrJurisdiction</i>
<i>creator</i>	Heredada de DCTERMS. Indica el <i>dcterms:Agent</i> responsable de hacer un recurso
<i>educationLevel</i>	Heredada de DCTERMS. Indica el <i>dcterms:AgentClass</i> una clase de entidad, definida, en términos de progresión, a través de un contexto educativo o de formación, para la que se destina un recurso descrito
<i>eventOrganisation</i>	Vincula una <i>dcterms:Organisation</i> que actúa como agente en un <i>core:Event</i> . Superpropiedad: <i>core:notablyAssociatedWith</i>
<i>eventPerson</i>	Vincula una <i>Person</i> que actúa como agente en un <i>core:Event</i> . Superpropiedad: <i>core:notablyAssociatedWith</i>
<i>eventPlace</i>	Indica la ubicación de un <i>Event</i> , que relaciona con <i>core:Place</i> . Superpropiedad: <i>core:notablyAssociatedWith</i>
<i>eventTheme</i>	Relación un <i>core:Event</i> y un <i>core:Theme</i> , del que es un factor. Superpropiedad: <i>core:notablyAssociatedWith</i>
<i>extent</i>	Heredada de DCTERMS. Indica la <i>dcterms:SizeOrDuration</i> de un recurso
<i>format</i>	Heredada de DCTERMS. Indica la <i>dcterms:MediaTypeOrExtent</i> de un recurso
<i>gender</i>	Permite distinguir <i>core:Person</i> por sexos

Propiedad	Descripción
<i>instructionalMethod</i>	Heredada de DCTERMS. Indica el <i>dcterms:MethodOfInstruction</i> de un recurso, proceso utilizado para generar conocimiento, actitudes y habilidades
<i>language</i>	Heredada de DCTERMS. Indica el <i>dcterms:LinguisticSystem</i> de un recurso
<i>license</i>	Heredada de DCTERMS. Indica el <i>dcterms:license</i> de un recurso
<i>locator</i>	Heredada de CMS. Vincula <i>core:Thing</i> y <i>cms:Locator</i>
<i>mediator</i>	Heredada de DCTERMS, indica el <i>dcterms:PhysicalMedium</i> de un recurso
<i>medium</i>	Heredada de DCTERMS, indica el <i>dcterms:AgentClass</i> que media en el acceso a un recurso y para quién dicho recurso está destinado o es útil
<i>notablyAssociatedWith</i>	Vincula una cosa a cualquier otra. Subpropiedades: <i>core:eventOrganisation</i> , <i>core:eventPerson</i> , <i>core:eventPlace</i> , <i>core:eventTheme</i> , <i>core:occupation</i> , <i>core:placeOfBirth</i> , <i>core:placeOfDeath</i>
<i>occupation</i>	Relaciona una persona (<i>core:Person</i>) y el área temática en la que trabaja (<i>Theme</i>). Superpropiedad: <i>core:notablyAssociatedWith</i>
<i>placeOfBirth</i>	Relaciona <i>core:Person</i> y su lugar de nacimiento (<i>core:Place</i>). Superpropiedad: <i>core:notablyAssociatedWith</i>
<i>placeOfDeath</i>	Relaciona <i>core:Person</i> y su lugar de fallecimiento (<i>core:Place</i>). Superpropiedad: <i>core:notablyAssociatedWith</i>
<i>primaryTopic</i>	Heredada de la ontología BBC, indica que un concepto (<i>core:Thing</i>) es el tema principal de un documento web (<i>bbc:webDocument</i>), definido como página disponible desde una URL en el portal de la BBC o cualquier documento web al que la BBC apunte. Inversa: <i>core:primaryTopicOf</i>
<i>primaryTopicOf</i>	Indica que un concepto (<i>core:Thing</i>) es el tema principal de un documento web (<i>bbc:webDocument</i>). Inversa: <i>core:primaryTopic</i>
<i>provenance</i>	Heredada de DCTERMS. Indica la <i>dcterms:ProvenanceStatement</i> de un recurso
<i>publisher</i>	Heredada de DCTERMS. Indica el <i>dcterms:Agent</i> responsable de hacer que un recurso esté disponible
<i>rights</i>	Heredada de DCTERMS. Indica los <i>dcterms:RightsStament</i> de un recurso
<i>rightsHolder</i>	De DCTERMS, indica el <i>dcterms:Agent</i> propietario o gestor de los derechos sobre un recurso
<i>sameAs</i>	Indica que algo es lo mismo que algo. Es ligeramente más débil que <i>owl:sameAs</i> . Su fin es vincular identidades separadas de la misma cosa mientras se mantiene la separación entre las declaraciones originales de cada cosa. En la ontología relaciona <i>core:Thing</i> y <i>owl:Thing</i>
<i>spatial</i>	Heredada de DCTERMS. Indica la <i>dcterms:Location</i> de un recurso
<i>tag</i>	Heredada de CWORK. indica que un trabajo creativo es etiquetado con un concepto (<i>core:Thing</i>). Aparece en el gráfico de la ontología, no en el fichero de la misma
<i>temporal</i>	Heredada de DCTERMS. Indica la <i>dcterms:PeriodOfTime</i> de un recurso
<i>type</i>	Heredada de DCTERMS. Indica la naturaleza o género de un recurso <i>dcterms:Class</i>

Tabla 29: Relaciones de BBCCORE

Propiedades tipo dato de BBCCORE:

Propiedad	Descripción
<i>alternative</i>	Heredada de DCTERMS. Nombre alternativo para un recurso
<i>available</i>	Heredada de DCTERMS. Fecha a partir de la que el recurso está disponible
<i>bibliographicCitation</i>	Heredada de DCTERMS. Referencia bibliográfica de un recurso
<i>created</i>	Heredada de DCTERMS. Fecha de creación de un recurso
<i>date</i>	Heredada de DCTERMS. Punto o periodo de tiempo asociado con un evento en el ciclo de vida de un recurso
<i>dateAccepted</i>	Heredada de DCTERMS. Fecha de aceptación de un recurso
<i>dateCopyrighted</i>	Heredada de DCTERMS. Fecha de copyright
<i>dateOfBirth</i>	Permite indicar la fecha de nacimiento de una <i>Person</i>
<i>dateOfDeath</i>	Permite indicar la fecha de muerte de una <i>Person</i>
<i>dateSubmitted</i>	Heredada de DCTERMS. Fecha de entrega de un recurso
<i>disambiguationHint</i>	Una o dos palabras que facilitan la distinción entre entidades con <i>labels</i> similares
<i>identifier</i>	Heredada de DCTERMS. Referencia unívoca para un recurso en un contexto dado
<i>issued</i>	Heredada de DCTERMS. Fecha de emisión de un recurso
<i>knowsFor</i>	Breve resumen del porqué una <i>Person</i> es notablemente conocida
<i>label</i>	Etiqueta para una cosa. Subpropiedad: <i>preferredLabel</i>
<i>modified</i>	Heredada de DCTERMS. Fecha en que se modificó un recurso
<i>preferredLabel</i>	La etiqueta preferida por la BBC para una cosa. Superpropiedad: <i>label</i>
<i>shortLabel</i>	Indica la etiqueta corta para una cosa, requerida para presentar propósitos
<i>slug</i>	Palabras clave legibles por personas que deben ser utilizadas de forma programática
<i>title</i>	Heredada de DCTERMS. Nombre dado a un recurso
<i>valid</i>	Heredada de DCTERMS. Fecha de validación de un recurso

Tabla 30: Propiedades tipo dato de BBCCORE

Como ya se ha indicado, BBCCORE reutiliza otras ontologías:

Tipo información	Dominio	Enlace
<i>DCAM</i>	Metadatos Dublin Core	http://dublincore.org/documents/abstract-model/
<i>DCTERMS</i>	Documentos	http://dublincore.org/documents/2012/06/14/dcmi-terms/
<i>BBC</i>	Documentos web BBC	http://www.bbc.co.uk/ontologies/bbc
<i>CMS</i>	Sistema gestión contenidos BBC	http://www.bbc.co.uk/ontologies/cms
<i>CWORK</i>	Contenido creativo BBC	http://www.bbc.co.uk/ontologies/creativework
<i>PROVENANCE</i>	Gráficos BBC	http://www.bbc.co.uk/ontologies/provenance

Tabla 31: Vocabularios reutilizados en BBCCORE

SIMPLE NEWS AND PRESS ONTOLOGIES (SNAP)

Es un grupo de ontologías para el modelado del mundo real, pues éste es el propósito de las

noticias. SNAP modela recursos (*assets*), entidades del mundo (entidades simples) y eventos (entidades complejas). En SNAP, el vínculo entre entidades y recursos se realiza a través de etiquetas (*tags*) y los recursos son categorizados mediante esquemas de clasificación, como *IPTC Media Topic Codes*, *schema.org Vocabulary* o *Press Association Categorisation* (McGinnis et al., 2011). SNAP Identifier es la que permite mapear las URIs de los recursos a través de identificadores creados por alguna autoridad. Esta visión da lugar a seis (micro)ontologías interrelacionadas (ver Tabla 32).

Ontología	Prefijo	Dominio	URL
<i>SNAP Asset</i>	pna	Recursos	http://data.press.net/ontology/asset/
<i>SNAP Classification</i>	pnc	Sistemas de clasificación	http://data.press.net/ontology/classification/
<i>SNAP Event</i>	pne	Eventos	http://data.press.net/ontology/event/
<i>SNAP Identifier</i>	pni	Identificadores	http://data.press.net/ontology/identifier/
<i>SNAP Stuff</i>	pns	Entidades simples	http://data.press.net/ontology/stuff/
<i>SNAP Tag</i>	pnt	Etiquetas	http://data.press.net/ontology/tag/

Tabla 32: Ontologías que conforman SNAP ontologies

Las 6 ontologías se relacionan mediante clases, de la forma en que se detalla en la Figura 81, basada en descripciones y modelos de los vocabulario⁶⁹: *pna:Asset* es subclase de *pnc:Classifiable* y *pnt:Taggable*; *pns:Stuff* de *pnt:Tag* y *pni:Identifiable*; *pne:Event* de *pns:Stuff*. Las clases y relaciones de SNAP se describen con más detalle posteriormente.

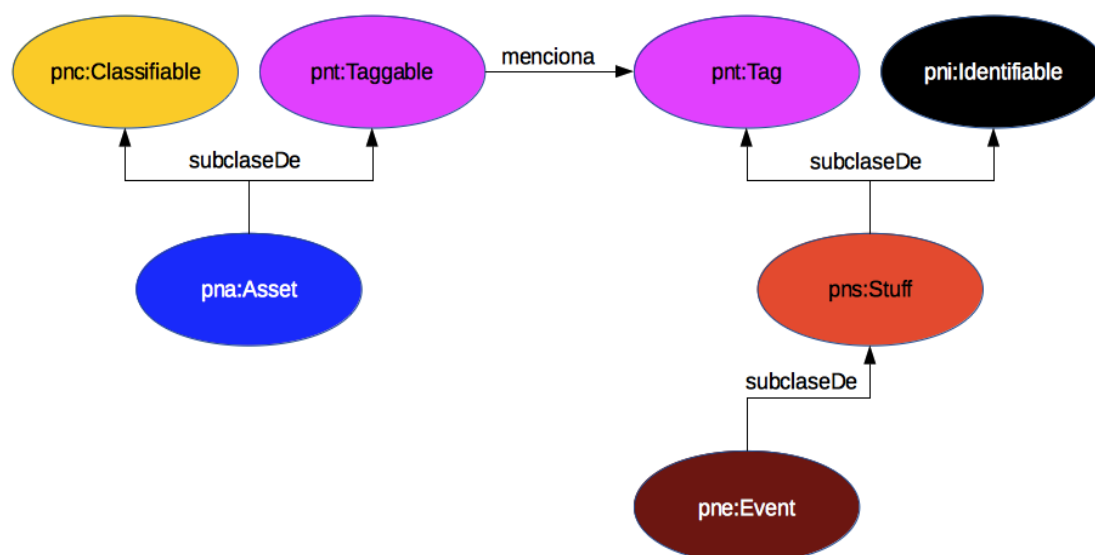


Figura 81: Interrelaciones entre las clases de las (micro)ontologías SNAP

Por otro lado, SNAP reutiliza elementos de otros vocabularios, que se indican en esta tabla:

⁶⁹ Fuente: <http://data.press.net/ontology/> (último acceso: 20/05/2017)

Vocabulario	Dominio	Enlace
DCMI	Recursos	http://dublincore.org/documents/dcmi-type-vocabulary/
EVENT	Eventos	http://motools.sourceforge.net/event/event.html
FOAF	Personas y organizaciones	http://xmlns.com/foaf/spec/
GEO	Espacios físicos	http://www.w3.org/2003/01/geo/
PROTON	Integración semántica de ontologías	http://ontotext.com/proton-ontology/
TIME	Espacios temporales	http://www.w3.org/TR/owl-time/

Tabla 33: Vocabularios reutilizados en el conjunto de las ontologías SNAP

A continuación se describe cada ontología SNAP. En general, las descripciones se basan en los archivos de las ontologías, que no siempre coinciden con las especificaciones.

SNAP ASSET ONTOLOGY (PNA)

Permite describir recursos de noticias (*Assets*), relaciones entre ellos y cómo pueden ser clasificados y anotados semánticamente. PNA está compuesta por 10 clases (7 importadas de PNC, DCMIType, FOAF y PNT); 4 relaciones y 11 propiedades tipo dato (4 importadas de DCTERMS) (ver Tabla 34, Tabla 35 y Tabla 36).

Clase	Descripción
<i>Classifiable</i>	Heredada de PNC. Permite clasificar un recurso utilizando alguna clasificación. Subclase: <i>pna:Asset</i> .
<i>Asset</i>	Permite clasificar un recurso según su tipología (vídeo, texto, imagen u otra). Subclases: <i>pna:Image</i> , <i>pna:Text</i> , <i>pna:Video</i> . Superclases: <i>pnc:Classifiable</i> , <i>pnt:Taggable</i>
<i>Image</i>	Identifica un recurso como imagen. Superclases: <i>pna:Asset</i> , <i>dcmitype:StillImage</i>
<i>Text</i>	Identifica un recurso como texto. Superclases: <i>pna:Asset</i> , <i>dcmitype:Text</i>
<i>Video</i>	Identifica un recurso como vídeo. Superclases: <i>pna:Asset</i> , <i>dcmitype:MovingImage</i>
<i>MovingImage</i>	Heredada de DCMIType. Tiene como subclase <i>pna:Video</i>
<i>Person</i>	Heredada de FOAF para identificar una persona. Sin relaciones
<i>StillImage</i>	Heredada de DCMIType. Subclase: <i>pna:Image</i>
<i>Taggable</i>	Heredada de PNT para etiquetar recursos. Subclase: <i>pna:Asset</i> .
<i>Text</i>	Importada de DCMIType. Subclase <i>pna:Text</i>

Tabla 34: Clases de la ontología (SNAP) PNA

Estas son las relaciones de (SNAP)PNA:

Propiedad	Descripción
<i>associatedAsset</i>	Asocia recursos. Subpropiedades: <i>pna:derivedFrom</i> , <i>pna:hasThumbnail</i> , <i>pna:hasTranscript</i>
<i>derivedFrom</i>	Asocia un recurso con otro recurso, del que deriva. Superpropiedad: <i>pna:associatedAsset</i>
<i>hasThumbnail</i>	Asocia una miniatura con un recurso tipo imagen. Superpropiedad: <i>pna:associatedAsset</i>
<i>hasTranscript</i>	Asocia un vídeo con un texto, que lo transcribe. Superpropiedad: <i>pna:associatedAsset</i>

Tabla 35: Relaciones de (SNAP) PNA

Finalmente, estas son sus propiedades tipo dato:

Propiedad	Descripción
<i>abstract</i>	Heredada de DCTERMS. Permite describir un recurso mediante un resumen. Subpropiedad: <i>pna:summary</i>
<i>byline</i>	Propiedad para indicar la firma o pie de autor de un recurso
<i>created</i>	Heredada de DCTERMS. Fecha de creación de un recurso. Subpropiedad: <i>pna:created</i>
<i>created</i>	Indica fecha y hora de creación de un recurso. Superpropiedad: <i>dcterm:created</i>
<i>embargoedUntil</i>	Permite indicar fecha y hora de embargo de un recurso a partir del cual se publica
<i>modified</i>	Heredada de DCTERMS. Indica fecha de modificación de un recurso. Subpropiedad: <i>pna:modified</i>
<i>modified</i>	Indica fecha y hora de modificación de un recurso. Superpropiedad: <i>dcterm:modified</i>
<i>published</i>	Permite indicar la fecha y hora de publicación de un recurso
<i>summary</i>	Permite resumir un recurso de noticias. Superpropiedad: <i>dcterm:abstract</i>
<i>title</i>	Heredada de DCTERMS para indicar el título de un recurso. Subpropiedad: <i>pna:title</i>
<i>title</i>	Propiedad para indicar el título de un recurso. Superpropiedad: <i>dcterms:title</i>

Tabla 36: Propiedades tipo dato de (SNAP) PNA

En el gráfico de la ontología se propone, mediante la propiedad de anotación *owl:seeAlso*, ver el vocabulario *Nature of a News Item* (*ninature*), de NewsCodes, de tipos de *media:Picture Item(s)*, *Text Item(s)*, *Video Item(s)*.

Se relaciona con PNC a través de *pnc:Classifiable* y con PNT a través de *pnt:Taggable*.

SNAP CLASSIFICATION ONTOLOGY (PNC)

Pensada para facilitar la clasificación de recursos mediante un vocabulario controlado. Las instancias del vocabulario heredan de *pnc:Classification*, permitiendo clasificar una instancia *pnc:Classifiable*. Compuesta por 3 clases (Tabla 37) y 3 relaciones (Tabla 38).

Clase	Descripción
<i>Classifiable</i>	Convierte un recurso en clasificable. Subclase <i>pna:Asset</i>
<i>Classification</i>	Las instancias de vocabulario deben ser de este tipo. Permite clasificar un <i>pnc:Classifiable</i>
<i>Person</i>	Importada de FOAF para identificar una persona

Tabla 37: Clases de la ontología (SNAP) PNC

Estas son las relaciones de (SNAP)PNC:

Propiedad	Descripción
<i>isClassifiedBy</i>	Relaciona <i>pnc:Classifiable</i> (elemento clasificable) y <i>pnc:Classification</i> . Subpropiedad: <i>dcterms:subject</i>

Propiedad	Descripción
<i>subject</i>	Heredada de DCTERMS para indicar el tema de un recurso. Subpropiedad: <i>isClassifiedBy</i>
<i>subClassificationOf</i>	Propiedad transitiva que relaciona una clasificación con una clasificación padre. Por tanto, relaciona <i>Classification</i> consigo misma

Tabla 38: Relaciones de (SNAP) PNC

En el gráfico de la ontología se muestra la propiedad de anotación *protons:transitiveOver*, que tiene como dominio *pnc:isClassifiedBy* y rango *pnc:subClassificationOn*. Dicha propiedad forma parte de la ontología PROTON (PROTo Ontology)⁷⁰.

Se relación con PNA a través de la clase *pnc:Classifiable*.

SNAP EVENT ONTOLOGY (PNE)

PNE está orientada a la descripción de eventos. *Event* es la única entidad compleja de SNAP. Un revisor de LOV, Bernard Vatant, decía en 2013 que es una extensión de *Event ontology* aplicada a las noticias. Esta ontología está compuesta por 5 clases, 3 relaciones y 8 propiedades tipo dato, algunos importadas de EVENT, FOAF, PNS y TIME. La especificación refleja varias relaciones y clases de EVENT y PNS.

Elemento	Descripción
<i>Event</i>	Heredada de EVENT. Permite identificar un evento. Subclase: <i>pne:Event</i>
<i>Event</i>	Permite identificar un evento. Superclases: <i>event:Event</i> y <i>pns:Stuff</i>
<i>Intangible</i>	Heredada de PNS para las cosas intangibles. En el gráfico de la ontología se indica como factor de un <i>Event</i> a través de <i>event:factor</i> . Superclase: <i>Stuff</i>
<i>Person</i>	Heredada de FOAF para identificar una persona. Aparece en el fichero de la ontología. Subclase: <i>pns:Person</i>
<i>Person</i>	Heredada de PNS, aparece en el gráfico de la ontología. Relacionada con <i>pne:Event</i> a través de <i>event:agent</i> . Superclases: <i>foaf:Person</i> , <i>pns:Tangible</i>
<i>Stuff</i>	Heredada de PNS, base para el resto de clases de la ontología PNS. Subclases: <i>pns:Event</i> , <i>pns:Tangible</i> , <i>pns:Intangible</i>
<i>Organization</i>	Heredada de PNS. Aparece en el gráfico de esta ontología. Relacionada con <i>pns:Event</i> a través de <i>event:agent</i> . Superclases: <i>foaf:Organization</i> , <i>pns:Tangible</i>
<i>Location</i>	Heredada de PNS. Aparece en el gráfico de esta ontología. Relacionada con <i>pns:Event</i> a través de <i>event:place</i> . Superclases: <i>geo:SpatialThing</i> y <i>pns:Tangible</i>
<i>Tangible</i>	Heredada de PNS, para las cosas tangibles. En el gráfico de la ontología, se muestra como factor de un <i>pns:Event</i> a través de <i>event:factor</i> . Subclases: <i>pns:Person</i> , <i>pns:Organization</i> , <i>pns:Location</i> . Superclase: <i>pns:Stuff</i>
<i>TemporalEntity</i>	Heredada de TIME, para definir entidades temporales. En el gráfico de la ontología es el momento en que ocurre un <i>pns:Event</i> a través de <i>event:time</i> . Subclases: <i>pns:Person</i> ,

⁷⁰ PROTo Ontology: Ontología de alto nivel desarrollada en el proyecto europeo SEKT. Actualmente su Qname ha sido sustituido por ptop: <http://ontotext.com/products/proton/> (último acceso: 20/05/2017)

Elemento	Descripción
	<i>pns:Organization, pns:Location</i> ; Superclase: <i>pns:Stuff</i>

Tabla 39: Clases de la ontología (SNAP) PNE

Relaciones de (SNAP)PNE:

Propiedad	Descripción
<i>notablyAssociatedWith</i>	Vincula de modo particular dos <i>pns:Stuff</i> . Subpropiedad: <i>pne:theme</i>
<i>theme</i>	Relaciona <i>pne:Event</i> y su tema, <i>pns:Intangible</i> . Superpropiedad: <i>notablyAssociatedWith</i>
<i>subEventOf</i>	Relaciona <i>pne:Event</i> consigo misma. Se trata de una propiedad transitiva

Tabla 40: Relaciones de la ontología (SNAP) PNE

Propiedades tipo dato de (SNAP)PNE:

Propiedad	Descripción
<i>abstract</i>	Heredada de DCTERMS. Describe un recurso con un resumen. Subpropiedad <i>pne:summary</i>
<i>comment</i>	Heredada de PNSn indica la descripción definitiva de un recurso. Subpropiedad <i>pne:summary</i>
<i>label</i>	Heredada de PNS, indica la etiqueta definitiva de 1 instancia <i>pns:Stuff</i> . Subpropiedad: <i>pne:title</i>
<i>summary</i>	Indica el resumen de un evento con una cadena de texto. Superpropiedades: <i>pns:comment</i> y <i>dcterms:abstract</i>
<i>title</i>	Heredada de DCTERMS para indicar el título de un recurso. Subpropiedad: <i>pne:title</i>
<i>title</i>	Propiedad para indicar el título de un recurso. Superpropiedades: <i>dcterms:title</i> y <i>pns:label</i>

Tabla 41: Propiedades tipo dato de (SNAP) PNE

SNAP IDENTIFIER ONTOLOGY (PNI)

Ontología para identificar recursos a partir de identificadores no basados en RDF. Este vocabulario permite mantener un mapeado entre los identificadores de recursos de RDF y sus equivalentes IDs en un dominio alternativo no basado en RDF. Compuesta por 4 clases (Tabla 42), 2 relaciones (Tabla 43) y una propiedad tipo dato (Tabla 44). Sólo importa una clase de FOAF y se relaciona con el resto de ontologías SNAP a través de *pns:Stuff*, una subclase de *pni:Identifiable*.

Clase	Descripción
<i>Authority</i>	Autoridad de un <i>pni:Identifier</i> , por ejemplo, el espacio de nombre dentro del cual un ID está contenido, expresado mediante un literal

Clase	Descripción
<i>Identifiable</i>	Instancia identificable. Subclase <i>pns:Stuff</i>
<i>Identifier</i>	ID de un recurso dado (<i>Identifiable</i>)
<i>Person</i>	Importada de FOAF para identificar una persona
<i>Stuff</i>	Base para el resto de clases de la ontología PNS. Superclase: <i>pni:Identifiable</i> . Aparece sólo en el gráfico de la ontología

Tabla 42: Clases de (SNAP) PNI

Las relaciones de (SNAP)PNI se recogen en la Tabla 43:

Propiedad	Descripción
<i>authority</i>	Relaciona un <i>pni:Identifier</i> y la autoridad que lo promueve, <i>pni:Authority</i>
<i>hasIdentifier</i>	Relaciona un recurso (<i>pni:Identifiable</i>) y su <i>pni:Identifier</i>

Tabla 43: Relaciones de (SNAP) PNI

Una única propiedad tipo dato forma parte de (SNAP)PNI:

Propiedad	Descripción
<i>value</i>	Permite indicar el valor, mediante un literal, de un <i>pni:Identifier</i>

Tabla 44: Propiedad tipo dato de (SNAP) PNI

SNAP STUFF ONTOLOGY (PNS)

PNS modela entidades del mundo real, distinguiendo entidades tangibles (personas, localizaciones y organizaciones) e intangibles (conceptos abstractos, como fumar, feminismo o amor). PNS se relaciona con FOAF, TIME, GEO, PNA y PNI y está compuesta por 18 clases (Tabla 45), 7 relaciones (Tabla 46) y 8 propiedades tipo dato (Tabla 47).

Clase	Descripción
<i>Asset</i>	Heredada de PNA. Hace referencia a un recurso
<i>Image</i>	Heredada de PNA. Identifica un recurso como imagen
<i>Identifiable</i>	Heredada de PNI. Instancia identificable. Subclase: <i>pns:Stuff</i>
<i>Intangible</i>	Base para las cosas intangibles. Superclase: <i>pns:Stuff</i>
<i>Instant</i>	Heredada de TIME. Permite indicar un instante
<i>Location</i>	Clase base para indicar una localización. Superclases: <i>geo:SpatialThing</i> y <i>pns:Tangible</i>
<i>Organization</i>	Importada de FOAF para identificar una organización. Superclase: <i>foaf:Agent</i> . Subclase: <i>pns:Organization</i>
<i>Organization</i>	Base para las instancias de organizaciones. Superclases: <i>foaf:Organization</i> , <i>pns:Tangible</i>

Clase	Descripción
<i>Person</i>	Importada de FOAF para identificar una persona. Aparece sólo en el fichero de la ontología
<i>Person</i>	Clase base para las instancias de personas. Superclases: <i>foaf:Person</i> y <i>pns:Tangible</i>
<i>SpatialThing</i>	Heredada de GEO para definir una entidad espacial. Subclase: <i>pns:Location</i>
<i>Stuff</i>	Base para el resto de clases de la ontología PNS. Superclases: <i>pni:Identifiable</i> y <i>pnt:Tag</i> . Subclases: <i>pns:Tangible</i> y <i>pns:Intangible</i>
<i>Tag</i>	Heredada de PNT. Clase general para indicar una etiqueta. Subclase: <i>pns:Stuff</i>
<i>Tangible</i>	Base para las cosas tangibles. Superclase: <i>pns:Stuff</i> . Subclases: <i>pns:Location</i> , <i>pns:Organization</i> , <i>pns:Person</i>

Tabla 45: Clases de (SNAP) PNS

Las relaciones de (SNAP)PNS son las siguientes:

Propiedad	Descripción
<i>dateOfBirth</i>	Fecha de nacimiento de una persona. Vincula <i>pns:Person</i> y <i>time:Instant</i>
<i>dateOfDeath</i>	Fecha de fallecimiento de una persona. Vincula <i>pns:Person</i> y <i>time:Instant</i>
<i>describedBy</i>	Relaciona una entidad de dominio (<i>pns:Stuff</i>) con documentos que la describen
<i>hasAsset</i>	Asocia una entidad de dominio (<i>pns:Stuff</i>) y un <i>pna:Asset</i> . Subclase: <i>pns:hasImage</i> . Puede que aparezca en el gráfico como <i>preferredAsset</i>
<i>hasImage</i>	Asocia una entidad de dominio (<i>pns:Stuff</i>) y un recurso tipo imagen (<i>pna:Image</i>). Superclase: <i>pns:hasAsset</i> . Puede que aparezca en el gráfico como <i>preferredImage</i>
<i>Location</i>	Propiedad reflexiva
<i>notablyAssociationWith</i>	Relaciona un elemento de la clase <i>pns:Stuff</i> consigo misma
<i>placeOfBirth</i>	Relaciona una persona (<i>pns:Person</i>) con su lugar de nacimiento (<i>pns:Location</i>)

Tabla 46: Relaciones de (SNAP) PNS

Las propiedad tipo dato de (SNAP)PNS se especifican en la Tabla 47.

Propiedad	Descripción
<i>alias</i>	Indica el alias de un <i>pns:Stuff</i> . Superpropiedad: <i>rdf:label</i>
<i>comment</i>	Indica la descripción definitiva de una instancia de <i>pns:Stuff</i> . Superpropiedad: <i>rdfs:comment</i>
<i>label</i>	Indica el alias de un <i>pns:Stuff</i> . Superpropiedad: <i>rdf:label</i> . Subpropiedades: <i>pns:longName</i> , <i>pns:name</i> y <i>pns:shortName</i>
<i>longName</i>	Indica el nombre completo de un <i>foaf:Agent</i>
<i>shortName</i>	Indica el nombre corto de un <i>foaf:Agent</i>
<i>name</i>	Heredada de FOAF. Indica el nombre de algo. Subclase: <i>pns:name</i>
<i>name</i>	Indica nombre completo y definitivo de <i>foaf:Agent</i> . Superpropiedades: <i>foaf:name</i> y <i>pns:label</i>

Tabla 47: Propiedades tipo dato de (SNAP) PNS

SNAP TAG ONTOLOGY (PNT)

PNT define relaciones para cosas etiquetables (*taggable things*) utilizando entidades de dominio (*stuff*) y eventos (*events*). Para ello, cuenta con 5 clases (Tabla 48) y 3 propiedades (Tabla 49). Se relaciona con el resto de ontologías SNAP mediante *pnt:Taggable* y su subclase *pna:Asset*, así como *pnt:Tag* y su subclase *pns:Stuff*.

Clase	Descripción
<i>Asset</i>	Heredada de PNA, hace referencia a un recurso. Superpropiedad: <i>pnt:Taggable</i>
<i>Person</i>	Importada de FOAF para identificar una persona
<i>Stuff</i>	Heredada de PNT. Base para el resto de clases de la ontología PNS. Superpropiedad: <i>pnt:Tag</i>
<i>Tag</i>	Clase genérica para referir una etiqueta. Proporciona una abstracción de interfaz (<i>interface abstraction</i>) para entidades de dominio. Subpropiedad: <i>pns:Stuff</i>
<i>Taggable</i>	Cosas que pueden ser etiquetadas. Define una interfaz que permite que los recursos sean etiquetados sin limitar qué clase de recursos son etiquetables. Subpropiedad: <i>pna:Asset</i>

Tabla 48: Clases de (SNAP) PNT

Las relaciones de (SNAP)PNS son las que se indican en la tabla que sigue:

Propiedad	Descripción
<i>about</i>	Relaciona un elemento etiquetable (<i>pnt:Taggable</i>) y una etiqueta (<i>pnt:Tag</i>), refinando la propiedad <i>pnt:isTaggedWith</i> . Esta relación es más fuerte que <i>pnt:mentions</i> . Superpropiedad: <i>pnt:isTaggedWith</i>
<i>mentions</i>	Relaciona un elemento etiquetable (<i>pnt:Taggable</i>) y una etiqueta (<i>pnt:Tag</i>), refinando la propiedad <i>pnt:isTaggedWith</i> . Superpropiedad: <i>pnt:isTaggedWith</i>
<i>isTaggedWith</i>	Relaciona <i>pnt:Taggable</i> y <i>pnt:Tag</i> . Subpropiedades: <i>pnt:about</i> y <i>pnt:mentions</i>

Tabla 49: Relaciones de (SNAP) PNT

POLITICS ONTOLOGY FOR WEB ENTITY RETRIEVAL (POWER)

Se trata de una ontología de procesos políticos, que define y describe los participantes en la actividad política a partir de roles en el ámbito político y sus relaciones. Esta herramienta es desarrollada teniendo en cuenta algunos de los retos que suponen para los sistemas de información la representación y descripción de cuestiones políticas en los medios de comunicación (Moreira et al., 2011). Entre estos destacan la multiplicidad de formas en que un sujeto político puede ser mencionado y la frecuencia con que cambian de rol en la escena política. Así, POWER especifica estructuras e instituciones de los poderes políticos, a cualquier nivel (local, nacional, supranacional) y geografía; relaciones entre políticos/as, organizaciones políticas e instituciones; elecciones, procesos por los que logran el mandato en una institución política, etc. El objetivo es describir los diferentes sistemas electorales a nivel local, nacional y supranacional. POWER es desarrollada en el marco del proyecto REACTION (*Retrieval, Extraction and Aggregation Computing Technology for Integrating and Organizing News*), de periodismo computacional y consta de las clases que

se indican en la Tabla 50.

Clase	Descripción
<i>Election</i>	Proceso electivo para un mandato. No aparece en la especificación
<i>ElectionType</i>	Tipo de elecciones. Valores posibles: <i>Internal, Legislatives, Presidentials, Regional</i> . No aparece en la especificación
<i>EntityName</i>	Nombre por el que una entidad es conocida. Su fin es separar nombre y concepto
<i>Mandate</i>	Oficio de <i>Politician</i> en una <i>Organization</i> durante un periodo de tiempo delimitado. Sus miembros (instancias) tienen la forma de <i>cargo_fecha_codigoInstancia</i>
<i>MandateAttainment</i>	Forma en que se consigue un mandato. Valores: <i>Election, Appointment, Inherency</i>
<i>NameType</i>	Tipo de nombre de <i>EntityName</i> . Valores: <i>Acronym, Birthday, MediaName, Nick-Name</i>
<i>Organization</i>	Organización. Una organización tiene un <i>PoliticalScope</i> . Superclase: <i>PoliticalEntity</i> . Subclases: <i>PoliticalAssociation</i> y <i>PoliticalInstitution</i>
<i>PoliticalAfiliation</i>	Afiliación política de un individuo a una <i>PoliticalAssociation</i>
<i>PoliticalAssociation</i>	Asociación política, como un partido o una organización no gubernamental (ONG)
<i>PoliticalEntity</i>	Entidad política. Subclases: <i>Politician</i> y <i>Organization</i>
<i>PoliticalInstitution</i>	Institución que representa un poder político, como el Parlamento o el Gobierno de un país. Superclase: <i>Organization</i>
<i>PoliticalPower</i>	Tipo de poder de <i>PoliticalInstitution</i> . Valores: <i>Executive, Legislative, Moderator</i>
<i>PoliticalScope</i>	Alcance (geográfico) de intervención de una <i>PoliticalInstitution</i> . Valores posibles: <i>European, International, National, Regional</i>
<i>Politician</i>	Individuo político. Superclase: <i>PoliticalEntity</i>
<i>RunnersList</i>	Grupo de candidatos que participan en una <i>Election</i> , puede tener el respaldo de una <i>PoliticalAssociation</i> . No aparece en la especificación
<i>Statement</i>	Clase RDF, sin descripción. No aparece en la especificación
<i>webPage</i>	Página web de una <i>PoliticalEntity</i>

Tabla 50: Clases de POWER

Algunas clases de POWER no aparecen en la especificación: *Election*, *ElectionType* y *RunnerList* aunque sí en el fichero de la ontología. *Statement*, clase RDF, tampoco aparece en la especificación, y carece de descripción en la ontología.

Las relaciones de POWER se describen en la Tabla 51.

Propiedad	Descripción
<i>affiliatedTo</i>	Relaciona <i>Politician</i> y la <i>PoliticalAssociation</i> a la que está afiliado. La relación no es directa, tiene como dominio <i>PoliticalAffiliation</i> , que permite indicar cuánto dura dicha afiliación
<i>closeMatch</i>	Heredada de SKOS. Indica cuasisinonimia entre clases en ontologías que se alinean
<i>created</i>	Relación heredada de DCTERMS para indicar la fecha de creación de algo
<i>creator</i>	Relación heredada de DCTERMS. Indica el agente responsable de la inserción de un hecho
<i>exactMatch</i>	Heredada de SKOS. Indica la sinonimia entre dos clases en ontologías que se alinean

Propiedad	Descripción
<i>forMandate</i>	Una <i>Election</i> es para uno o varios <i>Mandate</i>
<i>hasAffiliation</i>	Relaciona <i>PoliticalAssociation</i> y <i>Politician</i> afiliado/a, usando <i>PoliticalAffiliation</i> como clase intermedia
<i>hasNature</i>	Relaciona <i>Mandate</i> y su naturaleza, a través de la clase <i>MandateAttainment</i>
<i>hasPower</i>	Relaciona <i>PoliticalInstitution</i> y el tipo de poder que ostenta, <i>PoliticalPower</i>
<i>hasScope</i>	Relaciona <i>Organization</i> y su alcance de intervención, <i>PoliticalScope</i>
<i>hasType</i>	Relaciona una <i>EntityName</i> y el tipo de nombre, a través de la clase <i>NameType</i>
<i>hasWeb</i>	Relaciona una <i>PoliticalEntity</i> y su <i>webPage</i>
<i>isCandidate</i>	Relaciona un <i>Politician</i> y la <i>RunnersList</i> en la que concurre en unas elecciones
<i>isComposedBy</i>	Utilizada para definir una <i>PoliticalAssociation</i> compuesta por otras <i>PoliticalAssociation</i> , en casos de coaliciones. Propiedad inversa: <i>isPartOf</i>
<i>isHeadCandidate</i>	Indica la relación entre un <i>Politician</i> y una <i>RunnersList</i> de la que es cabeza de lista
<i>isPartOf</i>	Indica la pertenencia de una <i>PoliticalAssociation</i> a otra <i>PoliticalAssociation</i> en casos de coalición. Propiedad inversa: <i>isComposedBy</i>
<i>isReferredBy</i>	Relaciona una <i>PoliticalEntity</i> y el <i>EntityName</i> que la refiere. Inversa: <i>refers</i>
<i>mandateIn</i>	Mandato (<i>Mandate</i>) de un <i>Politician</i> en una <i>PoliticalOrganization</i>
<i>refers</i>	Relaciona <i>EntityName</i> y la <i>PoliticalEntity</i> a la se refiere. Inversa: <i>isReferredBy</i>
<i>runsIn</i>	Una <i>RunnersList</i> participa en una <i>Election</i> concreta
<i>servesMandate</i>	Un <i>Politician</i> sirve en un <i>Mandate</i>
<i>withEndorsementOf</i>	Un <i>Politician</i> es avalado por una <i>PoliticalAssociation</i>
<i>source</i>	Heredada de DCTERMS, indica la fuente de los datos

Tabla 51: Relaciones de POWER

Algunas propiedades (externas) aparecen definidas en la especificación de la ontología, aunque después no han sido implementadas en el fichero de POWER. Es el caso de *dcterm:created*, *dcterm:creator*, *dcterm:source*, *skos:closeMatch* y *skos:exactMatch*.

Finalmente, estas son las propiedades tipo dato de POWER.

Propiedad	Descripción
<i>area</i>	Área de intervención política: economía, defensa, educación, etc.
<i>category</i>	Categoría que la página web de la <i>PoliticalEntity</i>
<i>foundedIn</i>	Fecha en que fue fundada una <i>PoliticalAssociation</i>
<i>from</i>	Indica fecha de inicio de un intervalo en una <i>PoliticalAffiliation</i> , <i>Organization</i> o <i>Mandate</i>
<i>geoNetID</i>	Alineamiento con la ontología geográfica GeoNet para una <i>Organization</i>
<i>id</i>	ID. Sin descripción. Aparece en el fichero de la ontología
<i>imageURL</i>	URL de una imagen de una clase <i>Politician</i>
<i>language</i>	Idioma. Aparece en la especificación de POWER
<i>name</i>	Valor para una <i>EntityName</i>

Propiedad	Descripción
<i>office</i>	Oficina en que se lleva a cabo un <i>Mandate</i>
<i>politicalWing</i>	Permite indicar la orientación política de una <i>PoliticalAssociation</i>
<i>term</i>	Término de oficina para una <i>Organization</i>
<i>to</i>	Indica fecha fin de un intervalo en una <i>PoliticalAffiliation</i> , <i>Organization</i> o <i>Mandate</i>
<i>url</i>	Permite indicar la URL de una <i>webPage</i>

Tabla 52: Propiedades tipo dato de POWER

POWER contenía 3590 políticos portugueses, 74 asociaciones políticas, 3043 nombres de entidades políticas, 5959 mandatos en instituciones políticas (Moreira et al., 2013). La Figura 82, de (Moreira et al., 2011), muestra el modelo de POWER con la instancia Politician, José Sócrates.

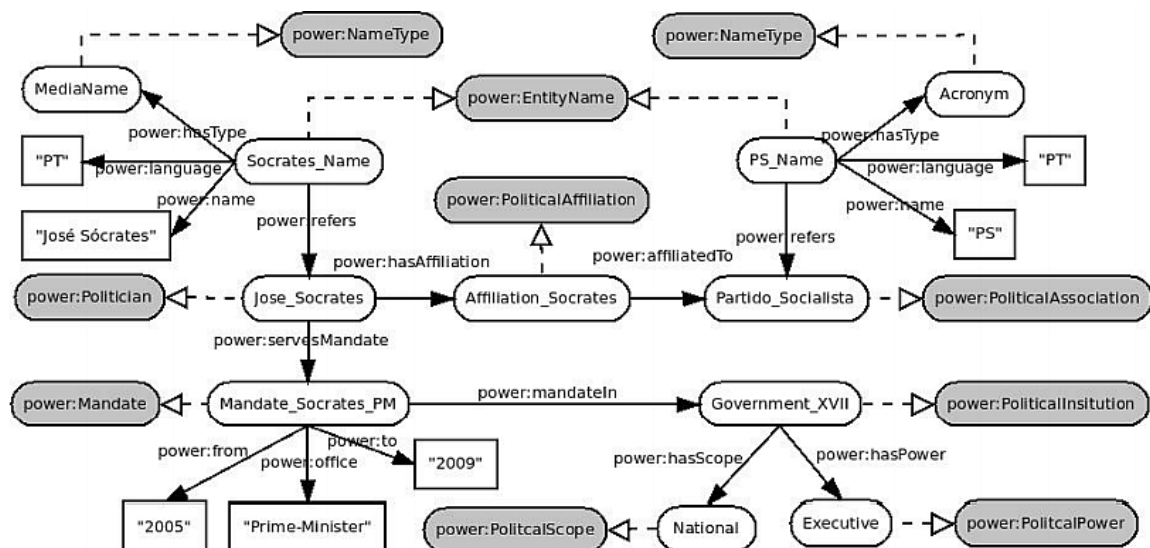


Figura 82: POWER aplicado a la instancia "José Sócrates" (Moreira et al., 2013)

Finalmente, en POWER se reutilizan elementos de las siguientes ontologías:

Tipo información	Dominio	Enlace
DCTERMS	Recursos	http://dublincore.org/documents/dcmi-type-vocabulary/
SKOS	Esquemas conceptuales	http://www.w3.org/2004/02/skos/

Tabla 53: Vocabularios reutilizados en POWER

THE POLITICAL ROLES ONTOLOGY (PROLES)

PROLES es una ontología en OWL2 que describe roles políticos de agentes y los relaciona con eventos. Además, permite establecer vínculos entre las descripciones y los documentos

de los que se extrae dicha información. Para ello reutiliza varias ontologías, especialmente de *The Publishing Roles Ontology* (PRO) (ShottonyPeroni, 2013). El punto de partida de PROLES es la importancia de las instituciones (*Corporate body*), personas (*Person*) y familia (*Family*) como entidades a describir en un registro de autoridad según la Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias (ISAAR(CPF)) (Daquino et al., 2014). Para su creación, los autores parten de cinco proposiciones: 1) Existe relación entre un agente y el rol que mantiene; 2) Cada rol es identificado unívocamente y es definido a través de una taxonomía de tipologías; 3) La relación agente-rol tiene un intervalo de tiempo y un contexto de aplicabilidad; 4) Esta relación puede ser vinculada a eventos que acontecen en un intervalo temporal y espacio geográfico concretos; 5) La relación entre el agente y el rol y la descripción de los eventos relacionados se basa en un meta-contexto de procedencia de los datos, compuesto por los documentos de los que se obtienen dichas relaciones y/o eventos y actores.

PROLES reutiliza varios vocabularios e incluye individuos específicos del ámbito político-económico. Dado el tamaño de PROLES, las clases, relaciones y atributos de vocabularios reutilizados se muestran en tablas independientes. Primero, se presentan todas las **clases** de la ontología (ver Tabla 54).

Clase	Descripción
<i>ParticipationWithPoliticalRole</i>	Situación específica en la que los agentes están implicados en relaciones políticas y participan en algún evento. La participación incluye a agentes, roles políticos en un tiempo y otros objetos, se restringe a un periodo de tiempo, se atestigua en una fuente y atribuye al editor de una afirmación
<i>Place</i>	Lugares en los que los agentes actúan con un rol político
<i>PoliticalRole</i>	Describe el rol político que puede tener una persona
<i>PoliticalRoleInTime</i>	Situación específica en la que los agentes están implicados en relaciones políticas, restringidas en intervalos de tiempo, atestiguadas en una fuente y atribuidas al editor de una afirmación. Superclase: <i>pro:RoleInTime</i>
<i>Source</i>	Fuentes que proporcionan información sobre roles de personas y contextos de las relaciones

Tabla 54: Clases específicas de PROLES

A continuación se describen las clases heredadas de otros vocabularios. En el caso de **FOAF**, son: *Agent*, *Document*, *Group*, *Image*, *OnlineAccount*, *Organization*, *Person*, *PersonProfileDocument* y *Project*. Ver Tabla 14 (página 286), este vocabulario también es reutilizado en esta investigación; Las clases heredadas de **TIME** son: *DayOfWeek*, *DateTimeDescription*, *DateTimeInterval*, *DurationDescription*, *January*, *Instant*, *Interval*, *ProperInterval*, *TemporalEntity*, *TemporalUnit* y *Year*. Se describen en la Tabla 18 (página 290), ya que este vocabulario es reutilizado en esta investigación. El resto de clases heredadas son las que se indican en la Tabla 55:

Clase	Descripción
Clases heredadas de PRO	
<i>PublishingRole</i>	Rol que un agente puede tener en relación con un proceso de publicación
<i>Role</i>	Rol que una persona podría tener. Subclases: <i>PoliticalRole</i> , <i>PublishingRole</i>
<i>RoleInTime</i>	Situación particular que describe el rol que un agente podría tener, restringiendo a un intervalo de tiempo. Describe siempre: un rol particular R, un agente A que tiene un rol R, un intervalo de tiempo T en el que el agente A tiene el rol R (opcional) y una o más entidades que definen el contexto en el que A tiene dicho R. Superclase: <i>tvc:ValueInTime</i> , Subclase: <i>PoliticalRoleInTime</i>
Clases heredadas de NARYPARTICIPATION	
<i>NaryParticipation</i>	Situación en la que la participación implica un <i>naryparticipation:TimeInterval</i> , <i>naryparticipation:Event</i> , <i>naryparticipation:Object</i> . Superclase: <i>situacion:Situation</i>
Clases heredadas de PARTICIPATION	
<i>Event</i>	Cualquier proceso físico, social o mental, evento o estado
<i>Object</i>	Cualquier objeto físico, social o mental, o sustancia
Clases heredadas de SITUATION	
<i>Situation</i>	Visión de un conjunto de entidades, contexto situacional. Subclases: <i>NaryParticipation</i> , <i>TimeIndexedSituation</i>
Clases heredadas de TIME INDEXED SITUATION	
<i>TimeIndexedSituation</i>	Una situación indexada de manera explícita, en algún momento, para, al menos, una entidad. Superclase: <i>Situation</i> . Subclase: <i>tvc:ValueInTime</i> . Superclase: <i>situacion:Situation</i>
Clases heredadas de TIME INTERVAL	
<i>TimeInterval</i>	Cualquier región en un espacio dimensional para representar el tiempo. Se usa cuando se quiere hablar de tiempo como objeto con características. Equivalente a <i>time:Interval</i>
Clases heredadas de TVC	
<i>ValueInTime</i>	Define un tipo particular de situación indexada de tiempo que representa un eje, vinculando la entidad que tenga un valor particular, el valor en sí y las extensiones temporales y contextuales de la que la relación entidad-valor depende. Superclase: <i>tsit:TimeIndexedSituation</i> . Subclase: <i>pro:RoleInTime</i>
Clases heredadas de TZONT	
<i>TimeZone</i>	Para indicar la zona horaria en que tiene lugar una entidad temporal (<i>TemporalEntity</i>)

Tabla 55: Clases de PROLES heredadas de varios vocabularios

Por lo que respecta a las relaciones de PROLES, se muestran en la Tabla 56:

Propiedad	Descripción
<i>hasRelatedAgent</i>	Relaciona <i>foaf:Agent</i> y <i>foaf:Place</i> con <i>foaf:Organization</i> o <i>foaf:Person</i>
<i>hasRelatedSource</i>	Relaciona <i>foaf:Agent</i> y <i>foaf:place</i> con una <i>Source</i> , que aporta información sobre estos
<i>includesAgentWithPoliticalrole</i>	Relaciona <i>politicalroles:ParticipationWithPoliticalRole</i> y un <i>foaf:Agent</i> que <i>pro:holdsRoleInTime</i> <i>politicalroles:PoliticalRoleInTime</i> . Superpropiedad: <i>naryparticipation:ParticipationIncludes</i> . Inversa:

Propiedad	Descripción
	<i>politicalroles:isAgentInParticipationWithPoliticalRole</i>
<i>includesPoliticalRoleInTime</i>	Relaciona una situación particular, <i>politicalroles:ParticipationWithPoliticalRole</i> a una situación de tiempo indexada, representando los roles políticos de los agentes implicados (<i>politicalroles:PoliticalRoleInTime</i>). Superpropiedad: <i>naryparticipation:participationIncludes</i> . Inversa: <i>politicalroles:isPoliticalRoleInTimeInParticipationWithPoliticalRole</i>
<i>isAgentInParticipationWithPoliticalRole</i>	Relaciona un <i>foaf:Agent</i> que <i>pro:holdsRoleInTime</i> algún <i>PoliticalRoleInTime</i> y una situación en la que participa con dicho rol, esto es, <i>ParticipationWithPoliticalRole</i> . Superpropiedad: <i>pro:isIncludedInParticipation</i> . Inversa: <i>includesAgentWithPoliticalRole</i>
<i>isPlaceContextFor</i>	Relaciona un <i>Place</i> con el rol para el que este lugar proporciona el contexto. Superpropiedad: <i>isRelatedToRoleInTime</i> , Inversa: <i>relatesToPlace</i>
<i>isPoliticalRoleInTimeInParticipationWithPoliticalRole</i>	Relaciona una situación de tiempo indexada representando los roles políticos de los agentes implicados (<i>PoliticalRoleInTime</i>) y una situación particular, <i>ParticipationWithPoliticalRole</i> . Superpropiedad: <i>isIncludedInParticipation</i> . Inversa: <i>includesPoliticalRoleInTime</i> .
<i>relatesToPlace</i>	Relaciona un rol para el que un <i>Place</i> proporciona un contexto y dicha clase. Superpropiedad: <i>relatesTo</i> . Inversa: <i>isPlaceContextFor</i>

Tabla 56: Relaciones específicas de PROLES

Las relaciones heredadas de FOAF son: *currentProject*, *depiction*, *depicts*, *fundedBy*, *homepage*, *interest*, *is primaryTopicOf*, *knows*, *logo*, *made*, *maker*, *mbox*, *member*, *page*, *pastProject*, *phone*, *primaryTopic*, *publications*, *schoolHomepage*, *thumbnail*, *tipjar*, *topic*, *topic_interest*, *weblog*, *workInfoHomepage* y *workplaceHomepage* y se describen en la Tabla 15 (página 288); Por lo que respecta a las relaciones heredadas de TIME (*after*, *before*, *dayOfWeek*, *hasBeginning*, *hasDateTimeDescription*, *hasDurationDescription*, *hasEnd*, *img* o *image*, *inDateTime*, *inside*, *intervalAfter*, *intervalBefore*, *intervalContains*, *intervalDuring*, *intervalEquals*, *intervalFinishedBy*, *intervalFinishes*, *intervalMeets*, *intervalOverlappedBy*, *intervalOverlaps*, *intervalStartedBy*, *intervalStarts*, *timeZone* y *unitType*), se indican en la Tabla 19 (página 291). El resto de propiedades heredadas son las que se indican en la Tabla 57.

Propiedad	Descripción
Relaciones heredadas de PRO	
<i>holdsRoleInTime</i>	Relaciona un <i>foaf:Agent</i> y <i>pro:RoleInTime</i> . Superpropiedad: <i>tv:hasValue</i> . Inversa: <i>pro:isRoleHeldBy</i>
<i>isDocumentContextFor</i>	Relaciona <i>foaf:Document</i> y el <i>pro:RoleInTime</i> para el que dicho documento proporciona un contexto. Superpropiedad: <i>pro:isRelatedToRoleInTime</i> . Inversa: <i>pro:relatesToDocument</i>
<i>isOrganizationContext</i>	Relaciona <i>foaf:Organization</i> y el rol para el que dicha organización proporciona un contexto. Superpropiedad: <i>pro:isRelatedToRoleInTime</i> . Inversa: <i>pro:relatesToOrganization</i>
<i>isRelatedToRoleInTime</i>	Relaciona una entidad al <i>pro:RoleInTime</i> para el que proporciona un contexto. Superpropiedad: <i>pro:hasTimeIndexedSetting</i> . Subpropiedades: <i>pro:isDocumentContextFor</i> , <i>pro:isPersonContextFor</i> , <i>proles:isPlaceContextFor</i> ; <i>pro:isOrganizationContextFor</i> . Inversa: <i>pro:relatesTo</i>
<i>isRoleHeldBy</i>	Relaciona un <i>pro:RoleInTime</i> de un <i>foaf:Agent</i> y dicho agente. Superpropiedad: <i>pro:forEntity</i> . Inversa: <i>pro:holdsRoleInTime</i>
<i>isRoleInTime</i>	Relaciona rol y la situación de un agente en dicho rol (<i>pro:RoleInTime</i>). Superpropiedad: <i>pro:hasTimeIndexedSetting</i> . Inversa: <i>pro:withRole</i>
<i>relatesTo</i>	Relaciona una situación de tiempo indexada (dominio: <i>pro:RoleInTime</i>) a una entidad <i>Thing</i> , representando el contexto para dicha situación. Superpropiedad: <i>tv:withinContext</i> . Subpropiedad: <i>pro:relatesToDocument</i> , <i>pro:relatesToOrganization</i> , <i>pro:relatesToPerson</i> , <i>proles:relatesToPlace</i>
<i>relatesToDocument</i>	Relaciona un <i>pro:RoleInTime</i> para el que un <i>foaf:Document</i> proporciona un contexto y dicha clase. Superpropiedad: <i>relatesTo</i> , Inversa: <i>isDocumentContextFor</i>
<i>relatesToOrganization</i>	Relaciona <i>pro:RoleInTime</i> para el que <i>foaf:Organization</i> proporciona un contexto y dicha clase. Superpropiedad: <i>relatesTo</i> , Inversa: <i>isOrganizationContextFor</i>
<i>relatesToPerson</i>	Relaciona <i>pro:RoleInTime</i> para el que un <i>foaf:Person</i> proporciona un contexto y dicha clase. Superpropiedad: <i>relatesTo</i> , Inversa: <i>isPersonContextFor</i>
<i>withRole</i>	Relaciona el valor que toma un agente en un momento determinado (<i>pro:RoleInTime</i>) con la definición del tipo de rol, especificado como instancia de <i>pro:Role</i> o una subclase. Superpropiedad: <i>pro:withValue</i>
Relaciones heredadas de NARY PARTICIPATION	
<i>isIncludedInParticipation</i>	Relaciona <i>Thing</i> y una situación <i>naryparticipation:NaryParticipation</i> . Superpropiedad: <i>situation:hasSetting</i>
<i>participationIncludes</i>	Relaciona una <i>Situation</i> y <i>Thing</i> . Superpropiedad: <i>situation:isSettingFor</i> . Subpropiedad: <i>situation:includesAgentWithPoliticalRole</i> y <i>situation:includesPoliticalRoleInTime</i>
Relaciones heredadas de PARTICIPATION	
<i>hasParticipant</i>	Permite relacionar un <i>participation:Event</i> y un <i>participation:Object</i> . Inversa: <i>participation:isParticipantIn</i>
<i>isParticipantIn</i>	Relaciona un <i>participation:Object</i> y el <i>participation:Event</i> en el que participan. Inversa: <i>participation:hasParticipant</i>

Propiedad	Descripción
Relaciones son heredadas de PROV-O:	
<i>had primary source</i>	Relaciona situaciones particulares (<i>PoliticalRoleInTime</i> y <i>ParticipationWithPoliticalRole</i>) con la fuente que las describe (<i>Source</i>)
<i>is person context for</i>	Propiedad que relaciona una persona al rol que tiene en un contexto. Superpropiedad: <i>pro:isRelatedToRoleInTime</i> . Inversa: <i>pro:relatesToPerson</i>
<i>was attributed to</i>	Relaciona una situación particular (<i>PoliticalRoleInTime</i> o <i>ParticipationWithPoliticalRole</i>) a un <i>foaf:Agent</i> responsable de las afirmaciones sobre los roles de dicho agente
Relaciones son heredadas de SITUATION	
<i>has setting</i>	Relaciona entidades (<i>Thing</i>) y una <i>situation:Situation</i> . Subpropiedades: <i>tisit:hasTimeIndexedSetting</i> , <i>naryparticipation:isIncludedInParticipation</i> y <i>tisit:isTimeIndexFor</i> . Inversa: <i>isSettingFor</i>
<i>isSettingFor</i>	Relaciona una <i>situation:Situation</i> y <i>Thing</i> . Subpropiedades: <i>tvc:atTime</i> , <i>tisit:atTime</i> , <i>tisit:forEntity</i> , <i>naryparticipation:participationIncludes</i>
Relaciones son heredadas de TISIT	
<i>atTime</i>	Periodo (rango: <i>tisit:TimeInterval</i>) durante el que un rol es mantenido o una contribución es realizada por un agente (dominio: <i>tisit:TimeIndexedSituation</i>). Subpropiedad: <i>tisit:isSettingFor</i> . Inversa: <i>tisit:isTimeIndexFor</i>
<i>forEntity</i>	Relaciona una <i>tisit:TimeIndexedSituation</i> definida para una entidad <i>Thing</i> . Superpropiedad: <i>tisit:isSettingFor</i> . Subpropiedades: <i>pro:isRoleHeldBy</i> , <i>tvc:withValue</i> , <i>tvc:withinContext</i> . Inversa: <i>tisit:hasTimeIndexedSetting</i>
<i>hasTimeIndexedSetting</i>	Relaciona <i>Thing</i> y una <i>tisit:TimeIndexedSituation</i> definida para dicha <i>Thing</i> . Superpropiedad: <i>situation:hasSetting</i> . Subpropiedad: <i>tvc:hasValue</i> , <i>pro:isRelatedToRoleInTime</i> , <i>pro:isRoleIn</i> . Inversa: <i>tisit:forEntity</i>
<i>isTimeIndexFor</i>	Relaciona un <i>tisit:TimeInterval</i> y una situación <i>tisit:TimeIndexedSituation</i> . Superpropiedad: <i>situation:hasSetting</i> . Inversa: <i>tisit:atTime</i>
Relaciones son heredadas de TVC	
<i>atTime</i>	Indica el periodo de tiempo, expresado a través de <i>tvc:ValueInTime</i> en que ocurre una situación (<i>time:TemporalEntity</i>). Superpropiedad: <i>tvc:isSettingFor</i>
<i>hasValue</i>	Relaciona una entidad <i>Thing</i> y una situación particular <i>tvc:ValueInTime</i> . Superpropiedad: <i>tisit:hasTimeIndexedSetting</i> . Subpropiedad: <i>pro:holdsRoleInTime</i>
<i>withinContext</i>	Vincula al contexto específico en el que en el cual el hecho de la entidad teniendo un valor es relevante. Dominio: <i>tvc:ValueInTime</i> , rango: <i>Thing</i> . Superpropiedad: <i>tvc:forEntity</i> . Subpropiedad: <i>pro:relatesTo</i>
<i>withValue</i>	Relaciona el valor (<i>tvc:valueInTime</i>) de una entidad que participa en una situación con dicha entidad (<i>Thing</i>). Superpropiedad: <i>tvc:forEntity</i> . Subpropiedad: <i>pro:withRole</i>

Tabla 57: Propiedades tipo dato de PROLES heredadas de varios vocabularios

Finalmente, PROLES sólo contiene propiedades tipo dato heredadas, de tres vocabularios. FOAF, TIME y TIME INTERVAL. Las relaciones heredadas de FOAF (*aimChatID*, *birthday*, *familyName*, *firstName*, *geekcode*, *gender*, *givenName*, *jabberId*, *myersBriggs*, *name*, *nick*, *plan*, *sha1*, *surname*, *title* y *yahooChatID*) se describen en la Tabla 16. Las

relaciones heredadas de TIME (*day, dayOfYear, days, hour, hours, inXSDDateTime, minute, minutes, month, months, second, seconds, week, weeks, xsdDateTime, year y years*) se recogen en la Tabla 20. Las heredadas de TIME INTERVAL se indican en la Tabla 58.

Propiedad	Descripción
<i>hasIntervalDate</i>	Codifica valores en formato <i>xsd:date</i> para un <i>timeinterval:TimeInterval</i> . Admite varios tipos de datos y formatos, como <i>xsd:gYear, xsd:dateTime</i> . Subpropiedades: <i>timeinterval:hasIntervalEndDate</i> y <i>timeinterval:hasIntervalStartDate</i>
<i>hasintervalEndDate</i>	Fecha final de <i>timeinterval:TimeInterval</i> . Superpropiedad: <i>timeinterval:hasIntervalDate</i>
<i>hasintervalStartDate</i>	Fecha inicial de un <i>timeinterval:TimeInterval</i> . Superpropiedad: <i>timeinterval:hasIntervalDate</i>

Tabla 58: Propiedades tipo dato heredadas de TIME INTERVAL

Además de los anteriores, PROLES incluye varias instancias, de entre las que destacan las de dominio político: *anarchist, enemy, friend, influent person, institutional role, leader, militant, prisoner, refugee, sympathiser*. El resto proceden de PRO, con perfiles propios del ámbito de la publicación (*archivist, author, author's agent, biographer, compiler, critic, editor, editor in chief, executive editor, ghost, writer, ghost editor, illustrator, librarian, peer review, printer, producer, proof reader, reviewer, senior editor, series editor, translator, etc.*) y de TIME (*Friday, Monday, Saturday, Sunday, Thursday, Tuesday, unitDay, unitHour, unitMinute, unitMonth, unitSecond, unitWeek, unitYear, Wednesday*).

En la Figura 83 se muestra la ontología, con tres capas, A) las relaciones entre agentes y roles; B) atribuciones de roles y eventos; C) procedencia de la información.

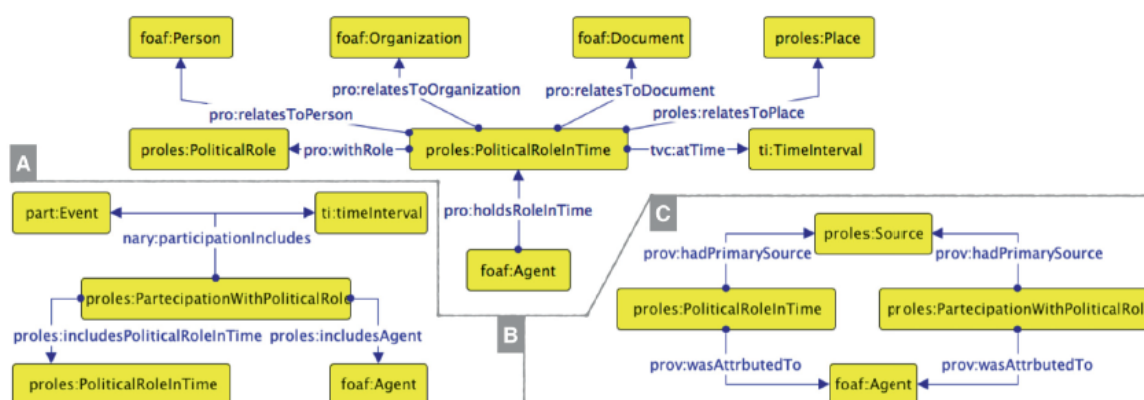


Figura 83: Modelo de la ontología Proles. Fuente: (Daquino et al., 2014)

Esta ontología, reutiliza elementos de otros vocabularios (ver Tabla 59), varios de los cuales han sido descritos anteriormente (FOAF, TIME, TIMEZONE, DC).

Tipo información	Dominio	Enlace
<i>DCTERMS</i>	Recursos	http://dublincore.org/documents/dcmi-terms/
<i>FOAF</i>	Agentes	http://www.foaf-project.org/
<i>NARY PARTICIPATION</i>	Eventos y sus participantes	http://ontologydesignpatterns.org/wiki/Submissions:Nary_Participation
<i>PARTICIPATION</i>	Participación de objetos en eventos	http://www.ontologydesignpatterns.org/ctp/owl/participation.owl
<i>PRO</i>	Roles de agentes	http://purl.org/spar/pro
<i>PROV-O</i>	Procedencia de información	http://www.w3.org/TR/prov-o/
<i>SITUATION</i>	Contextos o situaciones y cosas contextualizadas	http://www.ontologydesignpatterns.org/ctp/owl/situation.owl
<i>TIME</i>	Espacios temporales	http://www.w3.org/TR/owl-time/
<i>TIME INTERVAL</i>	Intervalos de tiempo	http://www.ontologydesignpatterns.org/ctp/owl/timeinterval.owl
<i>TIMEZONE</i>	Zonas horarias	http://www.w3.org/2006/timezone
<i>TISIT (Time-Indexed Situation Ontology)</i>	Situaciones temporales indexadas	http://www.ontologydesignpatterns.org/ctp/owl/timeindexedsituation.owl
<i>TVC (Time-Indexed Value in Context ontology)</i>	Entidades de valor concreto en contexto determinado	http://www.essepuntato.it/2012/04/tvc

Tabla 59: Vocabularios reutilizados en PROLES

A.1. HERRAMIENTAS REUTILIZADAS POR LOS VOCABULARIOS

La Tabla 60 resume qué herramientas han reutilizado los vocabularios seleccionados para el mapeo con ONA y ONAPE, de acuerdo con los criterios expuestos en 4.4.2. SELECCIÓN DE VOCABULARIOS.

La columna VOCABULARIO hace referencia a los vocabularios seleccionados, descritos en el apartado anterior, el encabezado de la tabla muestra las herramientas reutilizadas por estos, de acuerdo con la siguiente leyenda: A) BBC; B) BIBO; C) CMS; D) CWORK; E) DBPEDIA; F) DC; G) DNS; H) GOOD RELATIONS; I) NARY; J) PARTICIPATION; K) PRO; L) PROTON; M) PROV; N) PROVENANCE; O) SIOC; P) SITUATION; Q) TIME INTERVAL; R) TIMEZONE; S) TISIC; T) TVC; U) TZONT; V) WORDNET. Las columnas sombreadas corresponden a vocabularios seleccionados para el mapeo con ONA y ONAPE, también reutilizados por estos.

	VOCABULARIO	A	B	C	D	E	F	2	G	3	4	6	H	I	J	K	L	M	N	1	O	P	5	Q	R	S	T	U	V
1	SCHEMA		X				X	X			X		X								X								X
2	DUL							X	X																				
3	EVENT						X				X	X										X							X
4	FOAF						X					X									X								
5	TIME																												X
6	GEO																												
7	GEOP																												
8	BBCORE	X		X	X		X																						
9	PROLES						X				X			X	X	X		X				X	X	X	X	X	X	X	
10	POWER						X																						
11	SNAP						X			X	X	X				X							X						
	TOTAL	1	1	1	1	1	7	1	1	1	4	3	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	2	1

Tabla 60: Herramientas reutilizadas por los vocabularios a mapear con ONA y ONAPE

Como se observa, los vocabularios más reutilizados son Dublin Core (en cualquiera de sus variantes: DCTERMS, DCMI, etc.), FOAF, TIME y GEO.

A.2. ENLACES DE INTERÉS DE LOS VOCABULARIOS

Finalmente, la Tabla 61 recoge enlaces de interés de los vocabularios analizados siendo (a) especificación; (b) página principal del vocabulario; (c) url del archivo del vocabulario y (d) otros enlaces. Entre paréntesis, se indica la última fecha de actualización conocida.

VOCAB.		ENLACES
1	SCHEMA	(a) http://schema.org/ (2015) (b) http://schema.org/ (2015) (c) http://topbraid.org/schema/schema.rdf (2015) (d) Creación de código HTML con schema.org: http://schema-creator.org/ (2015); Página anterior: http://schema.rdfs.org/ (2015); Structured Data Testing Tool: https://developers.google.com/structured-data/testing-tool ; Schema Creator: http://schema-creator.org/ ; Thing Generator: http://www.microdatagenerator.com/thing-generators/
2	DUL	(b) http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite (2010) (c) http://www.ontologydesignpatterns.org/ont/dul/DUL.owl (2014)
3	EVENT	(a) http://motools.sourceforge.net/event/event.html (2007) (b) http://motools.sf.net/event/event.n3 (2007)
4	FOAF	(a) http://xmlns.com/foaf/spec/ (2014) (b) http://www.foaf-project.org/ [2014] (c) http://xmlns.com/foaf/spec/index.rdf [2014] (d) FOAF-a-matic: http://www.ldodds.com/foaf/foaf-a-matic.es.html
5	TIME	(b) http://www.isi.edu/~hobbs/owl-time.html [2007] (c) http://www.isi.edu/~hobbs/damlttime/time-entry.owl (2006)
6	GEO	(b) http://www.w3.org/2003/01/geo/wgs84_pos# [2009]
7	GEOP	(c) http://aims.fao.org/aos/geopolitical.owl [2015] (d) Datos instanciados en la ontología: http://www.fao.org/countryprofiles/iso3list/en/ ; Datos instanciados y estructura de la ontología: http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource/
8	BBCORE	(b) http://www.bbc.co.uk/ontologies/coreconcepts (2015) (c) http://www.bbc.co.uk/ontologies/coreconcepts/1.12.ttl (2015) (d) Ontologías de la BBC: http://www.bbc.co.uk/ontologies (2015)
9	PROLES	(a) http://www.essepuntato.it/lode/http://www.essepuntato.it/2013/10/politicalroles (c) http://eelst.cs.unibo.it/apps/LODE/source?url=http://www.essepuntato.it/2013/10/politicalroles
10	POWER	(a) http://dmir.inesc-id.pt/power/ [2011] (b) http://dmir.inesc-id.pt/project/POWER-PT_01_in_English [2011] (c) http://dmir.inesc-id.pt/project-media/images/d/d3/PowerBaseline.zip [2010] (d) Proyecto REACTIONS: http://dmir.inesc-id.pt/project/Reaction (2014)
11	SNAP	(b) pna: http://data.press.net/ontology/asset/ (2012); pnc: http://data.press.net/ontology/classification/ (2011); pne: http://data.press.net/ontology/event/ (2012); pni: http://data.press.net/ontology/identifier/ (2012); pns: http://data.press.net/ontology/stuff/ (2012); pnt: http://data.press.net/ontology/tag/ (2011) (d) Conjunto de ontologías SNAP: http://data.press.net/ontology/ [2012]; Imagen de interrelación entre ontologías SNAP: http://data.press.net/ontology/images/big-ontology-picture.png [2012]

Tabla 61: Enlaces de interés de los vocabularios a mapear con ONA y ONAPE

B. ESPECIFICACIÓN DE ONA

B1. INTRODUCCIÓN A LA ESPECIFICACIÓN

En este anexo se recoge la especificación de **ONA** (ONtología de Acciones). A diferencia de otras ontologías, ésta no cubre todos los aspectos relacionados con datos de gestión, pero puede ser utilizada para el modelado de documentos en cualquier dominio.

Para determinar el **propósito** de la ontología, uno de los métodos para determinar el alcance una ontología es plantear un listado de preguntas que debe responder (Shayji et al., 2011). En este sentido, se plantean las siguientes **cuestiones de competencia**: ¿qué evento se describe? ¿quién lo ha realizado? ¿quién acompaña al ejecutante de la acción? ¿qué consecuencias ha tenido dicha acción? ¿a quién a afectado? ¿cómo y con qué ha realizado esa acción? ¿cuál es la causa de esa acción? ¿y la motivación de quién la ejecuta? ¿es esta acción parte de un evento mayor? ¿cuándo ha ocurrido? ¿dónde ha tenido lugar? ¿de qué tipo de acción se trata? Un evento es algo muy complejo, donde se tienen que tener cuenta aspectos como el tiempo, el espacio, los objetos y las personas implicadas, así como las relaciones mereológicas, de causa y correlación entre ellos. Todos estos elementos responden a las cuestiones planteadas anteriormente.

Por lo que respecta al **léxico** de la ontología, no hay patrones o guías, si no que se deja a criterio del/de la ingeniero/a (Aguado-de-Cea et al., 2015). Incluso, dentro de la misma ontología pueden encontrarse diferencias. Se siguen las recomendaciones de estas autoras y la notación *CamelBack*, descrita en la página Error: no se encontró el origen de la referenciaError: no se encontró el origen de la referencia.

Esta especificación ha sido elaborada a partir de esta tesis y del fichero owl de la ONA. Para generar la documentación, el código fuente OWL de la ontología ha sido procesada a través de [LODE](#), Live OWL Documentation Environment, desarrollada por Silvio Peroni, descrita brevemente en el apartado 4.6. ESPECIFICACIONES DE LAS ONTOLOGÍAS.

IRI: <http://purl.org/umu/ona>

Version IRI: <http://purl.org/umu/ona>

B2. RESUMEN

ONA (Ontología de Acciones) es una ontología general desarrollada por María-José Baños-Moreno en el marco de la tesis doctoral "Propuesta de modelado de una ontología de dominio para la representación de acciones en política-economía".

Esta ontología ha sido modelada a partir de un componente teórico que desarrolla la Gramática de Casos de Fillmore (1967), García Gutiérrez (2014). Esta aproximación de descripción de noticias, de base lingüística, aplicable también a otro tipo de textos, analiza oraciones en función de las relaciones sintáctico-semánticas entre sus componentes, denominados "casos". García Gutiérrez, parte de las cuestiones de la Gramática de Casos de Fillmore para describir una noticia. Posteriormente, mediante procesos de indización humana y automática, se han añadido nuevos elementos, que vienen a completar aquellos definidos anteriormente. Para la definición de conceptos se utilizan fundamentalmente el Diccionario de la Real Academia de la Lengua Española (DLE) y Wikipedia.

Este producto puede ser útil para diferentes públicos, entre ellos, documentalistas, ya que contribuye a la representación de acciones descritas en documentos en general y en noticias en particular.

COMPONENTE TEÓRICO BÁSICO:

- Fillmore, Charles J. (1967). *The Case for Case*. In E. BachyR. Harms (Eds.), *Proceeding of the Texas Symposium on Linguistic Universals* (pp. 1–134). New York: Holt, Rinehart, and Winston. Retrieved from <http://eric.ed.gov/?id=ED019631>
- García Gutiérrez, A. (2014). *Análisis documental de noticias de prensa en sistemas de información factual*. *Revista Española de Documentación Científica*, 37(2), e046. <https://doi.org/10.3989/redc.2014.2.1094>

B3. ELEMENTOS DE ONA

- B3.1. CLASES
- B3.2. PROPIEDADES TIPO OBJETO
- B3.3. INDIVIDUOS DE ONA
- B3.4. DECLARACIÓN DE ESPACIOS DE NOMBRE EN ONA

B3.1. CLASES

Accion | Agente | Cosa | Cuatrimestre | DiaDeLaSemana | EspacioFisico | EspacioTemporal | Estacion | Grupo | Intangible | Intervalo | Mes | Objeto | Organizacion | Persona | Semestre | Situacion | Sujeto | Tangible | Trimestre

DESCRIPCIÓN DE LAS CLASES

Accion (IRI: <http://purl.org/umu/ona#Accion>)

Entidad básica que expresa el hecho que se produce en un EspacioTemporal y un EspacioFisico (conocidos o no), en el que puede participar, como mínimo, un Sujeto. Puede estar compuesto por Acciones menores.

está en el dominio de [esAntecedentePara](#), [esCausaDe](#), [esConsecuenciaDe](#), [tieneAntecedente](#), [tieneCausa](#), [tieneConsecuencia](#), [tieneEspacioFisico](#), [tieneEspacioTemporal](#), [tieneObjeto](#), [tieneResultado](#), [tieneSujeto](#)

está en el rango de [esAntecedentePara](#), [esCausaDe](#), [esConsecuenciaDe](#), [esEspacioFisicoDe](#), [esEspacioTemporalDe](#), [esObjetoDe](#), [esResultadoDe](#), [esSujetoDe](#), [tieneAntecedente](#), [tieneCausa](#), [tieneConsecuencia](#)

Agente (IRI: <http://purl.org/umu/ona#Agente>)

Ser que puede participar en una Accion, llevándola a cabo, acompañando a quien ejecuta dicha Accion o recibiendo sus efectos

tiene subclases [Grupo](#), [Organizacion](#), [Persona](#)

Cosa (IRI: <http://purl.org/umu/ona#Cosa>)

Todo lo que tiene entidad, tangible o intangible.

tiene subclases [Intangible](#), [Tangible](#)

está en el dominio de [esFinalidadDe](#), [esInstrumentoDe](#), [esModoDe](#), [esResultadoDe](#), [esSituacionDe](#)

está en el rango de [tieneFinalidad](#), [tieneInstrumento](#), [tieneModo](#), [tieneResultado](#), [tieneSituacion](#)

Cuatrimstre (IRI: <http://purl.org/umu/ona#Cuatrimstre>)

EspacioTemporal de cuatro meses. Habitualmente, para cuestiones estadísticas suele hablarse de tres cuatrimstres al año (primero, segundo y tercero).

tiene superclases [Intervalo](#)

DiaDeLaSemana (IRI: <http://purl.org/umu/ona#DiaDeLaSemana>)

Cada uno de los siete días de una semana.

es equivalente a { [Lunes](#) , [Martes](#) , [Miercoles](#) , [Jueves](#) , [Sabado](#) , [Domingo](#) , [Viernes](#) }

tiene superclases [Intervalo](#)

tiene miembros [Domingo](#), [Jueves](#), [Lunes](#), [Martes](#), [Miercoles](#), [Sabado](#), [Viernes](#)

EspacioFisico (IRI: <http://purl.org/umu/ona#EspacioFisico>)

Espacio físico o lugar en el que tiene lugar una Acción.

está en el dominio de [esEspacioFisicoDe](#)

está en el rango de [tieneEspacioFisico](#)

EspacioTemporal (IRI: <http://purl.org/umu/ona#EspacioTemporal>)

EspacioTemporal (fecha, periodo, tiempo) en el que tiene lugar una Accion

tiene subclases [Intervalo](#)

está en el dominio de [esEspacioTemporalDe](#)

está en el rango de [tieneEspacioTemporal](#)

Estacion (IRI: <http://purl.org/umu/ona#Estacion>)

Cada una de las cuatro partes o tiempos en que se divide el año de acuerdo a cambios meteorológicos, astronómicos, etc. Estos periodos son normalmente cuatro y duran aproximadamente tres meses y se denominan: primavera, verano, otoño e invierno

es equivalente a { [Primavera](#) , [Invierno](#) , [Verano](#) , [Otoño](#) }

tiene superclases [Intervalo](#)

tiene miembros [Invierno](#), [Otoño](#), [Primavera](#), [Verano](#)

Grupo (IRI: <http://purl.org/umu/ona#Grupo>)

Colección de Agentes individuales y/o Organizaciones.

tiene superclases [Agente](#)

Intangible (IRI: <http://purl.org/umu/ona#Intangible>)

Clase que engloba el conjunto de elementos intangibles, como pensamientos, sectores o ámbitos de actuación, roles, medidas, etc.

tiene superclases [Cosa](#)

tiene subclases [Situacion](#)

Intervalo (IRI: <http://purl.org/umu/ona#Intervalo>)

EspacioTemporal con extensión o duración.

tiene superclases [EspacioTemporal](#)

tiene subclases [Cuatrimestre](#), [DiaDeLaSemana](#), [Estacion](#), [Mes](#), [Semestre](#), [Trimestre](#)

Mes (IRI: <http://purl.org/umu/ona#Mes>)

Cada una de las doce partes en que se divide el año.

es equivalente a { [Julio](#) , [Diciembre](#) , [Mayo](#) , [Octubre](#) , [Enero](#) , [Febrero](#) , [Agosto](#) , [Abril](#) , [Noviembre](#) , [Junio](#) , [Marzo](#) , [Septiembre](#) }

tiene superclases [Intervalo](#)

tiene miembros [Abril](#), [Agosto](#), [Diciembre](#), [Enero](#), [Febrero](#), [Julio](#), [Junio](#), [Marzo](#), [Mayo](#), [Noviembre](#), [Octubre](#), [Septiembre](#)

Objeto (IRI: <http://purl.org/umu/ona#Objeto>)

Clase auxiliar que vincula un Agente o Cosa y la Accion de la que recibe los efectos de su ejecución. Esta clase “artificial” es utilizada para vincular todos los elementos relacionados con el Agente o Cosa en la recepción de los efectos de la Accion

está en el dominio de [esObjetoDe](#), [tieneReceptor](#)

está en el rango de [esReceptorDe](#), [tieneObjeto](#)

Organizacion (IRI: <http://purl.org/umu/ona#Organizacion>)

Tipo de Agente para organizaciones sociales, como empresas, sociedades, etc. Organizaciones y tipos de organización pueden ser denominadas con diversos nombres, como Agencia, Instituto, Comisión, Comité, Liga, Federación, Gabinete, Autoridad, Sociedad, Consejo etc., y no siempre tienen que ver con un ámbito o temática concreta. Por ejemplo, la Autoridad Aeroportuaria Británica y Autoridad de Servicios Financieros.

Las Organizaciones suelen tener un Alcance, AmbitoDeActuacion y un SectorDeActuacion.

tiene superclases [Agente](#)

Persona (IRI: <http://purl.org/umu/ona#Persona>)

Persona, independientemente de si está viva o muerta, si es real o imaginaria.

tiene superclases [Agente](#)

Semestre (IRI: <http://purl.org/umu/ona#Semestre>)

EspacioTemporal de seis meses. Habitualmente, para cuestiones estadísticas suele hablarse de dos semestres al año (primero y segundo).

tiene superclases [Intervalo](#)

Situacion (IRI: <http://purl.org/umu/ona#Situacion>)

Modo de ser o estar, estado o condición en que se encuentra alguien o algo.

tiene superclases [Intangible](#)

Sujeto (IRI: <http://purl.org/umu/ona#Sujeto>)

Clase auxiliar que vincula un Agente o Cosa y la Accion que ejecuta. Es utilizada para vincular todos los elementos relacionados con un Agente en la ejecución de una Accion

está en el dominio de [esSujetoDe](#), [tieneAcompañante](#), [tieneEjecutante](#), [tieneFinalidad](#), [tieneInstrumento](#), [tieneModo](#)

está en el rango de [esAcompañanteDe](#), [esEjecutanteDe](#), [esFinalidadDe](#), [esInstrumentoDe](#), [esModoDe](#), [tieneSujeto](#)

Tangible (IRI: <http://purl.org/umu/ona#Tangible>)

Clase que engloba el conjunto de elementos tangibles, como alimentos, productos o recursos.

tiene superclases [Cosa](#)

Trimestre (IRI: <http://purl.org/umu/ona#Trimestre>)

EspacioTemporal de tres meses. Habitualmente, para cuestiones estadísticas suele hablarse de cuatro trimestres al año (primero, segundo, tercero y cuarto).

tiene superclases [Intervalo](#)

B3.2. PROPIEDADES TIPO OBJETO

esAcompañanteDe | esAntecedentePara | esCausaDe | esConsecuenciaDe | esEjecutanteDe | esEspacioFisicoDe | esEspacioTemporalDe | esFinalidadDe | esInstrumentoDe | esModoDe | esObjetoDe | esParteDe | esReceptorDe | esResultadoDe | esSituacionDe | esSujetoDe | tieneAcompañante | tieneAntecedente | tieneCausa | tieneConsecuencia | tieneEjecutante | tieneEspacioFisico | tieneEspacioTemporal | tieneFinalidad | tieneInstrumento | tieneModo | tieneObjeto | tieneParte | tieneReceptor | tieneResultado | tieneSituacion | tieneSujeto

DESCRIPCIÓN DE LAS PROPIEDADES TIPO OBJETO

esAcompañanteDe (IRI: <http://purl.org/umu/ona#esAcompañanteDe>)

Relaciona Agente o Cosa y Sujeto

tiene dominio [Agente o Cosa](#); **tiene rango** [Sujeto](#)

es inversa de [tieneAcompañante](#)

esAntecedentePara (IRI: <http://purl.org/umu/ona#esAntedentePara>)

Propiedad que permite expresar la relación de antecendencia de una Acción sobre otra, cuando esa relación no es de causalidad ni de consecuencia.

tiene dominio [Accion](#); **tiene rango** [Accion](#)

es inversa de [tieneAntecedente](#)

esCausaDe (IRI: <http://purl.org/umu/ona#esCausaDe>)

Relaciona una Accion que esCausaDe otra Accion y dicha Accion

tiene dominio [Accion](#); **tiene rango** [Accion](#)

es inversa de [tieneCausa](#)

esConsecuenciaDe (IRI: <http://purl.org/umu/ona#esConsecuenciaDe>)

Relaciona la Accion consecuencia o efecto de una Accion y dicha Accion

tiene dominio [Accion](#); **tiene rango** [Accion](#)

es inversa de [tieneConsecuencia](#)

esEjecutanteDe (IRI: <http://purl.org/umu/ona#esEjecutanteDe>)

Relaciona el Agente o Cosa que ejecuta una Accion y el Sujeto a través del que la ejecuta

tiene dominio [Agente](#) o [Cosa](#); **tiene rango** [Sujeto](#)

es inversa de [tieneEjecutante](#)

esEspacioFisicoDe (IRI: <http://purl.org/umu/ona#esEspacioFisicoDe>)

Relaciona el EspacioFisico en que se produce una Accion y dicha Accion

tiene dominio [EspacioFisico](#); **tiene rango** [Accion](#)

es inversa de [tieneEspacioFisico](#)

esEspacioTemporalDe (IRI: <http://purl.org/umu/ona#esEspacioTemporalDe>)

Relaciona el EspacioTemporal en que se produce una Accion y dicha Accion

tiene dominio [EspacioTemporal](#); **tiene rango** [Accion](#)

es inversa de [tieneEspacioTemporal](#)

esFinalidadDe (IRI: <http://purl.org/umu/ona#esFinalidadDe>)

Relaciona la intención de un Agente en la ejecución de una Accion por medio de un Sujeto y ese Sujeto. Una Finalidad es entendida como Objetivo o Propósito de un Sujeto, a través de un Agente, por lo que esta relación tiene como dominio a un Agente y no a un Sujeto.

tiene dominio [Cosa](#); **tiene rango** [Sujeto](#)

es inversa de [tieneFinalidad](#)

esInstrumentoDe (IRI: <http://purl.org/umu/ona#esInstrumentoDe>)

Relaciona un Instrumento y el Agente que lo usa para ejecutar una Accion, por medio de la clase auxiliar Sujeto

tiene dominio [Cosa](#); **tiene rango** [Sujeto](#)

es inversa de [tieneInstrumento](#)

esModoDe (IRI: <http://purl.org/umu/ona#esModoDe>)

Relaciona una Cosa utilizada como Instrumento para la ejecución de una Accion y el Sujeto a través del que es utilizada

tiene dominio [Cosa](#); **tiene rango** [Sujeto](#)

es inversa de [tieneModo](#)

esObjetoDe (IRI: <http://purl.org/umu/ona#esObjetoDe>)

Relaciona el Objeto a través del que un Agente o Cosa recibe los efectos de una Accion y dicha Accion

tiene dominio [Objeto](#); **tiene rango** [Accion](#)

es inversa de [tieneObjeto](#)

esParteDe (IRI: <http://purl.org/umu/ona#esParteDe>)

Propiedad definida para expresar relaciones jerárquicas de tipo partonómico (parte-todo). Por ejemplo, entre una Acción que forma parte de otra más amplia.

es inversa de [tieneParte](#)

esReceptorDe (IRI: <http://purl.org/umu/ona#esReceptorDe>)

Relaciona el elemento, habitualmente un Agente o Cosa, que recibe los efectos una Accion y Objeto, clase auxiliar a través de la que los recibe. Puede interpretarse como esAfectadoPor

tiene dominio [Agente](#) o [Cosa](#); **tiene rango** [Objeto](#)

es inversa de [tieneReceptor](#)

esResultadoDe (IRI: <http://purl.org/umu/ona#esResultadoDe>)

Relaciona la Cosa producto de la ejecución de una Accion y dicha Accion

tiene dominio [Cosa](#); **tiene rango** [Accion](#)

es inversa de [tieneResultado](#)

esSituacionDe (IRI: <http://purl.org/umu/ona#esSituacionDe>)

Relaciona el estado en que se encuentra un Agente o Cosa, a través de un Sujeto u Objeto, en la ejecución o recepción de una Accion

tiene dominio [Cosa](#); **tiene rango** [Objeto](#) o [Sujeto](#)

es inversa de [tieneSituacion](#)

esSujetoDe (IRI: <http://purl.org/umu/ona#esSujetoDe>)

Relaciona Sujeto y la Accion que ejecuta

tiene dominio [Sujeto](#); **tiene rango** [Accion](#)

es inversa de [tieneSujeto](#)

tieneAcompañante (IRI: <http://purl.org/umu/ona#tieneAcompañante>)

Relaciona un Sujeto y el Agente o Cosa que lo acompaña, a través de la clase Sujeto, en la ejecución de una Accion

tiene dominio [Sujeto](#); **tiene rango** [Agente](#) o [Cosa](#)

es inversa de [esAcompañanteDe](#)

tieneAntecedente (IRI: <http://purl.org/umu/ona#tieneAntecedente>)

Propiedad que permite expresar la relación entre una Accion y otra que la antecede, cuando esa relación no es de causalidad ni de consecuencia.

tiene dominio [Accion](#); **tiene rango** [Accion](#)

es inversa de [esAntecedentePara](#)

tieneCausa (IRI: <http://purl.org/umu/ona#tieneCausa>)

Propiedad que relaciona una Accion y la Causa por la que ocurre

tiene dominio [Accion](#); **tiene rango** [Accion](#)

es inversa de [esCausaDe](#)

tieneConsecuencia (IRI: <http://purl.org/umu/ona#tieneConsecuencia>)

Propiedad que relaciona una Accion y la Consecuencia que produce

tiene dominio [Accion](#); **tiene rango** [Accion](#)

es inversa de [esConsecuenciaDe](#)

tieneEjecutante (IRI: <http://purl.org/umu/ona#tieneEjecutante>)

Relaciona un Sujeto y el Agente o Cosa que ejecuta una Accion a través del mismo

tiene dominio [Sujeto](#); **tiene rango** [Agente](#) o [Cosa](#)

es inversa de [esEjecutanteDe](#)

tieneEspacioFisico (IRI: <http://purl.org/umu/ona#tieneEspacioFisico>)

Relaciona una Accion y el EspacioFisico en que se produce

tiene dominio [Accion](#); **tiene rango** [EspacioFisico](#)

es inversa de [esEspacioFisicoDe](#)

tieneEspacioTemporal (IRI: <http://purl.org/umu/ona#tieneEspacioTemporal>)

Relaciona algo y el EspacioTemporal en que se produce. Por ejemplo, se puede establecer la relación entre una Acción y el EspacioTemporal en que ocurre.

tiene dominio [Accion](#); **tiene rango** [EspacioTemporal](#)

es inversa de [esEspacioTemporalDe](#)

tieneFinalidad (IRI: <http://purl.org/umu/ona#tieneFinalidad>)

Relaciona un Sujeto y la Finalidad de un Agente (a través de ese Sujeto) por la que ejecuta una Accion. Una Finalidad es entendida como objetivo, propósito o intención de un Agente a través de un Sujeto, por lo que esta relación tiene como dominio un Sujeto.

tiene dominio [Sujeto](#); **tiene rango** [Cosa](#)

es inversa de [esFinalidadDe](#)

tieneInstrumento (IRI: <http://purl.org/umu/ona#tieneInstrumento>)

Relaciona una cosa empleada como Instrumento por un Agente, a través de la clase Sujeto, para ejecutar una Accion

tiene dominio [Sujeto](#); **tiene rango** [Cosa](#)

es inversa de [esInstrumentoDe](#)

tieneModo (IRI: <http://purl.org/umu/ona#tieneModo>)

Relaciona el Modo en que un Agente, a través de la clase Sujeto, ejecuta una Accion, y ese Sujeto

tiene dominio [Sujeto](#); **tiene rango** [Cosa](#)

es inversa de [esModoDe](#)

tieneObjeto (IRI: <http://purl.org/umu/ona#tieneObjeto>)

Relaciona Accion y el Objeto que recibe los efectos de la ejecución de esa Accion

tiene dominio [Accion](#); **tiene rango** [Objeto](#)

es inversa de [esObjetoDe](#)

tieneParte (IRI: <http://purl.org/umu/ona#tieneParte>)

Propiedad definida para expresar relaciones jerárquicas de tipo partonómico (todo-parte). Por ejemplo, entre una Acción que forma parte de otra más amplia.

es inversa de [esParteDe](#)

tieneReceptor (IRI: <http://purl.org/umu/ona#tieneReceptor>)

Relaciona Objeto, como clase auxiliar, con el elemento, frecuentemente un Agente o Cosa, que recibe el efecto de una Accion. Puede interpretarse como sinónimo de expresiones que reflejan una idea similar, como tieneAfectado, afectaA, poneEnRiesgo, etc.

tiene dominio [Objeto](#); **tiene rango** [Agente](#) o [Cosa](#)

es inversa de [esReceptorDe](#)

tieneResultado (IRI: <http://purl.org/umu/ona#tieneResultado>)

Relaciona la Cosa resultante de la ejecución de una Accion y dicha Accion

tiene dominio [Objeto](#); **tiene rango** [Accion](#) o [Agente](#) o [Cosa](#) o [EspacioFisico](#) o [EspacioTemporal](#)

es inversa de [esReceptorDe](#)

tieneSituacion (IRI: <http://purl.org/umu/ona#tieneSituacion>)

Relaciona el Sujeto u Objeto y la Situación o Estado, definido a nivel de Cosa, en que se encuentra en la ejecución o recepción de una Accion

tiene dominio [Objeto](#) o [Sujeto](#); **tiene rango** [Cosa](#)

es inversa de [esSituacionDe](#)

tieneSujeto (IRI: <http://purl.org/umu/ona#tieneSujeto>)

Relaciona Accion y el Sujeto que la ejecuta

tiene dominio [Accion](#); **tiene rango** [Sujeto](#)

es inversa de [esSujetoDe](#)

B3.3. INDIVIDUOS DE ONA

Abril | Agosto | Diciembre | Domingo | Enero | Febrero | Invierno | Jueves | Julio | Junio | Lunes | Martes | Marzo | Mayo | Miercoles | Noviembre | Octubre | Otoño | Primavera | Sabado | Septiembre | Verano | Viernes

DESCRIPCIÓN DE LOS INDIVIDUOS DE ONA

Abril (IRI: <http://purl.org/umu/ona#Abril>)

Cuarto mes del año, que tiene 30 días.

pertenece a [Mes](#)

Agosto (IRI: <http://purl.org/umu/ona#Agosto>)

Octavo mes del año, que tiene 31 días.

pertenece a [Mes](#)

Diciembre (IRI: <http://purl.org/umu/ona#Diciembre>)

Duodécimo mes del año, que tiene 31 días.

pertenece a [Mes](#)

Domingo (IRI: <http://purl.org/umu/ona#Domingo>)

Séptimo día de la semana, que es festivo para el cristianismo y, en general, en el mundo occidental.

pertenece a [DiaDeLaSemana](#)

Enero (IRI: <http://purl.org/umu/ona#Enero>)

Primer mes del año, que tiene 31 días.

pertenece a [Mes](#)

Febrero (IRI: <http://purl.org/umu/ona#Febrero>)

Segundo mes del año, que en los comunes tiene 28 días y en los bisiestos 29.

pertenece a [Mes](#)

Invierno (IRI: <http://purl.org/umu/ona#Invierno>)

Época más fría del año, que en el hemisferio boreal corresponde a los meses de diciembre, enero y febrero, y en el austral a los de junio, julio y agosto.

pertenece a [Estacion](#)

Jueves (IRI: <http://purl.org/umu/ona#Jueves>)

Cuarto día de la semana.

pertenece a [DiaDeLaSemana](#)

Julio (IRI: <http://purl.org/umu/ona#Julio>)

Séptimo mes del año, que tiene 31 días.

pertenece a [Mes](#)

Junio (IRI: <http://purl.org/umu/ona#Junio>)

Sexto mes del año, que tiene 30 días.

pertenece a [Mes](#)

Lunes (IRI: <http://purl.org/umu/ona#Lunes>)

Primer día de la semana

pertenece a [DiaDeLaSemana](#)

Martes (IRI: <http://purl.org/umu/ona#Martes>)

Segundo día de la semana.

pertenece a [DiaDeLaSemana](#)

Marzo (IRI: <http://purl.org/umu/ona#Marzo>)

Tercer mes del año, que tiene 31 días.

pertenece a [Mes](#)

Mayo (IRI: <http://purl.org/umu/ona#Mayo>)

Quinto mes del año, que tiene 31 días.

pertenece a [Mes](#)

Miercoles (IRI: <http://purl.org/umu/ona#Miercoles>)

Tercer día de la semana.

pertenece a [DiaDeLaSemana](#)

Noviembre (IRI: <http://purl.org/umu/ona#Noviembre>)

Undécimo mes del año, que tiene 30 días.

pertenece a [Mes](#)

Octubre (IRI: <http://purl.org/umu/ona#Octubre>)

Décimo mes del año, que tiene 31 días.

pertenece a [Mes](#)

Otoño (IRI: <http://purl.org/umu/ona#Otoño>)

Época templada del año, que en el hemisferio boreal corresponde a los meses de septiembre, octubre y noviembre, y en el austral a los de marzo, abril y mayo

pertenece a [Estacion](#)

Primavera (IRI: <http://purl.org/umu/ona#Primavera>)

Época templada del año, que en el hemisferio boreal corresponde a los meses de marzo, abril y mayo, y en el austral a los de septiembre, octubre y noviembre.

pertenece a [Estacion](#)

Sabado (IRI: <http://purl.org/umu/ona#Sabado>)

Sexto día de la semana, que es festivo para el judaísmo y otras confesiones religiosas.

pertenece a [DiaDeLaSemana](#)

Septiembre (IRI: <http://purl.org/umu/ona#Septiembre>)

Noveno mes del año, que tiene 30 días. También se utiliza Setiembre.

pertenece a [Mes](#)

Verano (IRI: <http://purl.org/umu/ona#Verano>)

Época más calurosa del año, que en el hemisferio boreal corresponde a los meses de junio, julio y agosto, y en el austral a los de diciembre, enero y febrero.

pertenece a [Estacion](#)

Viernes (IRI: <http://purl.org/umu/ona#Viernes>)

Quinto día de la semana, que es festivo para el islamismo.

pertenece a [DiaDeLaSemana](#)

B3.4. DECLARACIÓN DE ESPACIOS DE NOMBRE EN ONA

espacio de nombre por defecto: <http://purl.org/umu/ona#>

owl: <http://www.w3.org/2002/07/owl#>

rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

rdfs: <http://www.w3.org/2000/01/rdf-schema#>

umu: <http://purl.org/umu/>

xsd: <http://www.w3.org/2001/XMLSchema#>

C. MARCADO SEMÁNTICO DE NOTICIAS DEL TEMA 1

En ese anexo se muestra el mercado semántico de las noticias que han sido empleadas como base para la segunda evaluación de ONA y ONAPE. La leyenda de colores se indica en el apartado 4.5.4. VALIDACIÓN POR ANOTACIÓN SEMÁNTICA.

NOTICIA 1: China media accuses Abe of dangerous politics (Los medios chinos acusan a Abe de política peligrosa). *The Asahi Shimbun*, 18/07/2013. Reuters, Beijing

Two of China's top newspapers accused Japanese Prime Minister Shinzo Abe on July 18 of dangerous politics that could threaten regional security, as Tokyo warned Beijing not to expand gas exploration in disputed waters of the East China Sea.

The People's Liberation Army Daily said Abe was trying to play the "China threat" angle, to win votes in July 21 elections, with a visit on July 17 to Japan's southern island of Ishigaki, near islets claimed by both China and Japan.

Territorial claims by Japan and China over the uninhabited islets and resource-rich waters in both the East China Sea and South China Sea rank as one of Asia's biggest security risks.

During the visit to Ishigaki island, Abe repeated Tokyo's stand that the nearby disputed Senkaku islands are inherent Japanese territory, adding that he has no intention of conceding even one step.

"This kind of 'drinking poison to slake ones thirst' not only threatens regional stability, it gives encouragement to Japan's 'turn to the right'," said the daily.

Abe wants to revise Japan's constitution, drafted by the United States after World War Two, to formalize the country's right to have a military. Critics say his plan could return Japan to a socially conservative, authoritarian past.

The People's Liberation Army Daily said Abe could not have chosen a worse time to visit Ishigaki, which lies some 160 km (100 miles) from the uninhabited islets the two nations contest.

"You cannot criticize a national leader for visiting his country's own territory but in a situation where the dispute over the Diaoyu Islands is continuing and the situation is complex and sensitive, Abe's actions are doubtless extremely dangerous and irresponsible," the paper, the official publication of China's military, said in a commentary.

The ruling Communist Party's official People's Daily warned that China would never allow

itself to be trampled on again, a reference to China's bitter memories of Japan's invasion of the country ahead of and during World War Two.

In a commentary published under the pen name "Zhong Sheng," or "voice of China," the newspaper said that Abe was looking for excuses to re-arm Japan and that the dispute with China was a convenient way of pushing this.

"The aim is to create tension and provoke incidents, to push Japan's military development," it said.

Patrol ships from both nations routinely shadow each other near the islands, raising concerns about an unintended clash.

On July 18, three Chinese surveillance vessels sailed into what Japan considers its territorial waters near the isles on what Beijing said was a routine patrol.

The Japan Coast Guard said the ships later left its territorial water but remain in the contiguous area.

The territorial dispute between China and Japan was further complicated on July 17 with news that Chinese state-run oil companies plan to develop seven new gas fields in the East China Sea, possibly siphoning gas from the seabed beneath waters claimed by Japan.

Beijing had slowed exploration in the energy-rich East China Sea but is now rapidly expanding its hunt for gas, a cheaper and cleaner energy to coal and oil imports.

Japanese Chief Cabinet Secretary Yoshihide Suga said his government had asked China about the plans, first reported by Reuters.

"If the Chinese side is to proceed unilaterally with development in the area over which there are conflicting claims, Japan would never accept it," Suga told a regular news conference on July 18.

NOTICIA 2: SURVEY: Only 48% aware of Japan's 'undisputed' ownership of Senkakus (Encuesta: sólo el 48% sabe de la propiedad "indiscutible" de Japón de las Senkakus). *The Asahi Shimbun*, 30/08/2013. The Asahi Shimbun

Only 48 percent of Japanese are aware of the central government's official position on the disputed Senkaku Islands, according to a Cabinet Office survey.

The survey, whose results were released Aug. 29, found that although 91.1 percent of the respondents know about the existence of the Senkakus, only 48 percent of them are aware of the government's position that the Senkaku Islands are under the valid control of Japan, and there exists no issue of territorial sovereignty to be resolved concerning the Senkaku Islands.

Located in the East China Sea, the dispute over the Senkaku Islands has strained bilateral relations between Japan and China since Tokyo's move to nationalize the Senkaku Islands, known as the Diaoyu Islands in China, in September 2012. The move set off vehement protests from China, which also claims the group of islets.

The survey was conducted between July 11-21 through interviews with 3,000 people age 20 or older across the nation. The Cabinet Office received valid responses from 1,801, or 60 percent.

According to the survey, the ratio of respondents who are "interested" in or "sort of interested" in the Senkakus issue was 73.7 percent.

Those "interested" or "sort of interested" were also asked to select one or more options that correspond to their concerns. Among the options, 62.6 percent picked "the basis of Japan's possession of the Senkaku Islands," and 54.9 percent picked "historical background."

Meanwhile, the ratio of respondents who are "not interested in" or "rather uninterested" in the Senkakus was 24.9 percent. Of these, 61.2 percent answered they "do not think the issue affects their daily lives that much."

NOTICIA 3: China asserts control over air zone despite US B-52 bombers (China afirma su control sobre la zona aérea a pesar de los bombarderos B-52 estadounidenses). *The Times of India*, 27/11/2013. PTI, Beijing

China on Wednesday asserted that it has enough "will and ability" to enforce its unilaterally declared new air defence zone over disputed islands in the East China Sea after two US B-52 bombers challenged Chinese authority over the controversial airspace.

Putting up a brave face, China's defence ministry said it "detected, identified and monitored" the flight of the giant long-range Stratofortress planes that flew the zone between two different times on Tuesday night.

"The Chinese government has the will and ability to defend our national sovereignty and security," foreign ministry spokesman Qin Gang said at a press briefing.

"We also have the ability to exercise effective control over the East Sea air defence identification zone," he said.

Defence ministry spokesman Geng Yansheng said the US aircraft skirted along the border of the zone and flew in the north-south direction, 200 kilometres east of Diaoyu Island.

China calls the disputed islands Diaoyus while Japan terms them as Senkakus.

Under the rules of the new air zone, all aircraft including the civilian flights have to report their flight plans to China, must maintain the two-way radio communications and respond in a timely and accurate manner to identification inquiries.

Those that do not comply can face "defensive emergency measures", Beijing said.

The US along with Japan refused to accept the zone announced by China over the disputed islands administered by Japan.

The Pentagon said it did not comply with Beijing's controversial demand for aircraft to file flight plans when traversing the East China Sea area.

US Colonel Steve Warren at the Pentagon said Washington had "conducted operations in the area of the Senkakus".

"We have continued to follow our normal procedures, which include not filing flight plans, not radioing ahead and not registering our frequencies," Warren said.

China's defence spokesman Geng said: "The Chinese military monitored the entire process, carried out identification in a timely manner, and ascertained the type of US aircraft.

"We need to stress that China will identify every aircraft flying in the air defence identification zone according to the country's announcement of aircraft identification rules for the air defence identification zone."

NOTICIA 4: China scrambles jets for first time in new air zone (China choca por primera vez en la nueva zona aérea). *The New York Times*, 29/11/2013. Jane Perlez, Beijing

China's Air Force scrambled jets on Friday and identified two American surveillance planes, and 10 Japanese aircraft, in the new air defense zone established by China over disputed islands in the East China Sea, the Chinese state news media said.

The scrambling of the jets was the first move announced by China that it was enforcing the air defense zone, which it established last weekend. China said at the time that foreign planes flying into the zone would be subject to military action if they did not present prior notification.

The American planes identified by the Chinese jets on Friday were a P-3 and an EP-3, said Col. Shen Jinke, a spokesman for China's Air Force, according to Xinhua, the state-run news agency.

The Chinese account said the 10 Japanese aircraft were of three types of military models. It named the E-767, an airborne warning and control system aircraft, the P-3 surveillance aircraft and the F-15 jet fighter, but did not say how many planes of each type.

A version of the EP-3 surveillance aircraft was involved in a major diplomatic incident between China and the United States in 2001 when an American EP-3 collided with a Chinese jet fighter over the South China Sea. The Chinese pilot was killed, and the American plane made an emergency landing on Hainan Island in southern China, an accident that badly damaged relations.

On Friday, asked for clarification on China's intentions in the new air zone, the spokesman at the Foreign Ministry, Qin Gang, said, "The Air Defense Identification Zone does not equal territorial airspace, and is not an expansion of a country's territorial airspace."

The spokesman also said, "Aircraft of all countries, including commercial aircraft, carrying out normal flight according to international law will not be affected."

The creation of the "air defense identification zone" by Beijing ratcheted up the tense relations between China and Japan, a treaty ally of the United States, and drew harsh criticism from Washington.

Many countries, including the United States and Japan, have air defense zones, but the coordinates of the Chinese zone overlap with parts of the Japanese zone, setting up what defense experts have called a provocative and dangerous situation in the airspace above the disputed islands.

The United States immediately challenged the new air zone and the threat of military action by dispatching two unarmed B-52 bombers into the zone on Monday without giving

notice to China. The Chinese took no action, saying only that they had monitored the B-52s.

Mr. Qin, the Foreign Ministry spokesman, brushed aside questions about Japanese criticism of China's air defense zone.

"Would the Japanese side tell other countries, does it have an A.D.I.Z.?" Mr. Qin said. "Has it negotiated with other countries while it established and enlarged its A.D.I.Z.? How large is its A.D.I.Z.?"

An American expert on air defense identification zones said Japanese aircraft would not be deterred from flying in the airspace above the disputed islands, known as the Diaoyu in China and the Senkaku in Japan. The islands are administered by Japan, but both countries claim them as sovereign territory.

The expert, Peter Dutton, the director of the China Maritime Studies Institute at the Naval War College in Newport, R.I., said that because Japan regards the airspace above the islands as national airspace, it would continue air patrols.

"Japan must continue to enforce its sovereignty or they could lose it to Chinese pressure,"

NOTICIA 5: ANALYSIS: With China still on his mind, Abe ends diplomatic work for 2013 (Análisis: Abe finaliza su año diplomático de 2013 con China aún en su mente). *The Asahi Shimbun*, 16/12/2013. Kotaro Ono y Atsushi Hiroshima

Prime Minister Shinzo Abe spent a year on the diplomatic front trying to bolster the economy, maintain Japan's image as a peace-loving country and keep China in check.

The results were mixed. And Tokyo's relations with Beijing and Seoul--two issues that have dominated headlines all year--will enter 2014 in a tattered state.

The prime minister wrapped up his diplomatic schedule for 2013 with the conclusion of the special session of Japan and the Association of Southeast Asian Nations in Tokyo on Dec. 15.

Abe met with the leaders of five ASEAN members, including Vietnam and Laos, on Dec. 15, and promised official development assistance (ODA) and yen loans worth about 170 billion yen (\$1.65 billion).

It is unclear if the ODA will prompt ASEAN members to award infrastructure contracts to Japanese companies. But the members did show that they were not ready to join Abe's plan of bringing Asian countries together to encircle China and its rising power.

For the joint statement issued Dec. 14 at the Japan-ASEAN summit, Japan wanted to add a clause criticizing China's new air defense identification zone (ADIZ) in the East China Sea that covers the Japanese-administered Senkaku Islands.

The ASEAN leaders rejected Japan's request.

At high-level talks between Japan and Russia in November, Tokyo also tried to bring Moscow on board against China.

But Moscow refused to take sides at the first "two-plus-two" talks, involving the foreign and defense chiefs of the two countries.

Since taking office in December 2012, Abe has gone overseas almost every month. He visited 25 countries and attended summits both at home and abroad with more than 110 leaders.

The summits did not include China.

Abe and South Korean President Park Geun-hye also did not hold formal talks in 2013 due to the dispute over the Takeshima islets, history issues related to World War II, and the hard-line stance of Park's administration.

PROMOTING TRADE

Deputy Chief Cabinet Secretary Katsunobu Kato, who accompanied Abe on his overseas trips in 2013, said he is confident that Abe's package of reforms known as "Abenomics" is increasing Japan's presence in foreign countries.

When Abe visited foreign countries in 2013, influential people in the business community, including Hiromasa Yonekura, chairman of Keidanren (Japan Business Federation), often accompanied him.

Abe has been particularly aggressive in trying to export Japan's nuclear power technology as a pillar of his "economic diplomacy."

The prime minister signed a nuclear energy agreement with the United Arab Emirates, the first since the Fukushima nuclear disaster started in 2011. He twice visited Turkey, and a corporate alliance led by a Japanese company won a formal order to build nuclear reactors in the country.

Japan has also been pushing exports of nuclear power plants to India, Vietnam, the Middle East and other countries.

But these aggressive moves have created a backlash from some in the Japanese public.

The Fukushima meltdowns heightened anti-nuclear sentiment around the nation, and critics say Abe should not be trying to sell such potentially dangerous technology. They also say the Fukushima disaster remains far from over.

The prime minister, however, has brushed aside such criticism.

"We now have a bad reputation for exporting nuclear power plants, but it should be highly evaluated 10 years later," he has been quoted as saying.

Yet not all is going according to plan for the nation's leader in this field.

The government planned to seek Diet approval of the nuclear energy agreement with Turkey during the extraordinary session in autumn. But the session was so focused on the state secrets protection law that the Turkey agreement was not discussed.

"We are in an awkward position because it was a bilateral commitment," a government official said.

CONCERNED ABOUT CHINA

Early in his administration, Abe was focused more on his ideals for Japan, and stressed that the nation will remain committed to peace.

To explain his vision of Japan's future role in the world, the prime minister used the slogan "proactive contribution to peace" in a speech before the United Nations General Assembly and in other occasions. He also considered visiting Israel and Palestine in October.

“Next year, we’d like him to visit countries that are important in terms of politics, in addition to economics,” said a senior official of the Foreign Ministry.

China, however, always seems to be in the back of Abe’s mind in diplomatic affairs.

In the latest flap over China’s internationally criticized ADIZ, the Chinese Foreign Ministry on Dec. 14 blasted Abe’s demand for Beijing to withdraw the zone. It also released a statement under the name of spokesman Hong Lei that said Japan’s effort to mislead world opinion is certain to fail.

Since March, Chinese President Xi Jinping and Premier Li Keqiang have visited a total of 22 countries, including those in Africa and Latin America, from a “resource diplomacy” perspective.

Abe plans to visit Africa in January, aiming to “drive a wedge” in China’s resource diplomacy, according to a senior official of his administration.

He has joined negotiations of the Trans-Pacific Partnership free trade arrangement, which does not include China, and has pursued efforts to support the Philippines and Vietnam, which both have territorial disputes with China in the South China Sea.

*

**