

## «Tendencias actuales de la documentación científica automatizada en la investigación educativa»

*por*  
*Luis Lizasoain*

### CUESTIONES GENERALES SOBRE DOCUMENTACIÓN CIENTÍFICA.

El actual desarrollo y evolución de las ciencias en general y de las llamadas sociales y humanas en particular conlleva una gran producción de documentación científica, que por su propia magnitud demanda la utilización de sistemas que permitan una eficaz y rápida recuperación y difusión de la misma para evitar la paradoja que pueda suponer el que dicho incremento de la información lleve implícito una cada vez mayor desconocimiento de los trabajos realizados en una disciplina cualquiera.

Se hace necesario pues disponer de sistemas que permitan el proceso de grandes masas de información y que además arbitren los canales de acceso y difusión para distribuirla allá donde se produce y necesita, evitando así la reduplicación de esfuerzos y facilitando la actualización de los conocimientos.

Evidentemente estas tareas son poco menos que irrealizables si no se cuenta con los recursos que proporcionan las nuevas tecnologías, especialmente la informática y la telemática.

En el ámbito de las Ciencias de la Educación también observamos como estas tecnologías están incidiendo prácticamente en todos los ámbitos de intervención educativa (proceso de enseñanza, aprendizaje, organización, administración, evaluación y medición, diagnóstico...) y en concreto en el de la investigación nos encontramos con importantes aplicaciones en lo que se refiere a análisis de datos, modelos de simulación, muestreo, etc.

En esta línea de potenciación de la infraestructura de la investigación educativa, la documentación automatizada ofrece una serie de ventajas que no poseen los sistemas tradicionales. Pero antes de pasar a examinar más detenidamente estas cuestiones veamos brevemente algunos conceptos básicos de dicha disciplina con el fin de aclarar y fijar la terminología usada.

Cuando se habla de un sistema documental automatizado (por ejemplo ERIC, uno de los más difundidos en educación) son varios los elementos que lo

integran, pero simplificando mucho el modelo, dos se nos aparecen como fundamentales: de una parte la base de datos propiamente dicha y de otra el Tesouro de la misma.

Obviamos aquí elementos como la red de ordenadores y terminales, el lógico necesario para el tratamiento y recuperación de la información, y todo lo referente a los distribuidores y comercializadores de la base de datos. Elementos imprescindibles pero para nuestros fines tienen en este momento menos relevancia.

Comenzando por la base de datos propiamente dicha, en primer lugar aclaremos que usamos el término desde el punto de vista documental, pues en informática tiene otras acepciones y matices. En este caso lo podríamos definir como *el conjunto de referencias documentales almacenadas y organizadas sistemáticamente para recuperación y difusión selectiva en función de los criterios del usuario*.

En cuanto a los tipos de bases de datos se suele hablar, entre otras, de bases de datos bibliográficas, textuales o factuales. Ejemplos de cada una de ellas serían la ya citada de ERIC, una de sentencias jurídicas o de legislación y, por último, las cotizaciones de bolsa de un periodo cualquiera.

En nuestro ámbito, si exceptuamos lo referente a la legislación educativa, las más usuales y útiles son las bibliográficas en las que las fuentes primarias son artículos, conferencias, ponencias, informes, memorias de investigación, tesis y en ocasiones libros. La información se almacena según el modelo bibliográfico por diferentes campos (autor, título, identificadores, descriptores).

El general una base de datos implica, disponer, además de la infraestructura tecnológica antedicha, de los fondos documentales a vaciar, de un equipo de documentalistas que realicen dicha labor y, por último de un lenguaje controlado y normalizado que permita la «indización» del contenido documental de forma unívoca, de manera que los términos que utilice el documentalista a la hora de definir los conceptos que aparecen en el documento sean los mismos que emplee el usuario cuando quiera recuperar la información.

Este lenguaje controlado y normalizado se plasma en el Thesaurus o Tesouro (que sobre su denominación no hay excesivo consenso).

La norma ISO 2788 define un Tesouro como sigue:

- según su función, es un instrumento de control terminológico usado por los «indizadores» o usuarios para traducir el lenguaje natural de los documentos a un lenguaje más controlado.
- según su estructura, es un vocabulario dinámico y controlado de términos relacionados semántica y genericamente que cubre un dominio específico del conocimiento.

Un Tesouro se compone de *descriptores* (términos o expresiones que designan, unívocamente los conceptos del campo cubierto por el tesouro), *no descriptores* (sinónimos o cuasi-sinónimos de los mismos) y *relaciones semánticas* (entre descriptores y no descriptores o entre descriptores).

Dichos descriptores se agrupan en conjuntos semánticos según diferentes criterios y entre ellos guardan relaciones jerárquicas (términos genéricos o específicos) y asociativas.

Como vemos el papel del tesoro es fundamental. Es el elemento que sirve de puente entre el documentalista y el usuario, permitiendo que la búsqueda y recuperación de la información sea lo más eficaz posible, o lo que es lo mismo, exhaustiva (se recuperan todos los documentos del tema elegido) y precisa (sólo los de ese tema).

Sin el tesoro las búsquedas serían muy poco eficaces especialmente en el campo de las ciencias sociales y humanas donde lo habitual es que los términos sean muy polisémicos a diferencia de las ciencias exactas y naturales en que tales casos suelen ser las excepciones. A cualquiera se nos pueden ocurrir infinidad de ejemplos sobre este particular, por lo que no creo preciso abundar más en ello.

Un tesoro además de su carácter funcional, expresa el estado actual del desarrollo terminológico de una disciplina en un momento y lugar determinados según el marco socio-cultural imperante, e igualmente fija los modelos de clasificación e inter-relación entre los diferentes campos de cada materia.

Para concluir digamos que el tesoro es el lenguaje en el que se construyen una o unas bases de datos y mediante el cuál se accede a las mismas; es la infraestructura lógico-lingüística de un sistema documental.

Pasemos ahora a examinar la situación en el campo de las Ciencias de la Educación.

## LA SITUACIÓN DE LAS CIENCIAS DE LA EDUCACIÓN

Si nos ceñimos a la información aportada por la obra *Bases de Datos del Mundo* (1), las bases de datos centradas en el área educativa en general o en alguna sub-área son las siguientes:

- «Child Abuse and Neglect» (prevención y tratamiento de abusos y negligencias en la educación infantil).
- «ERIC» (Ciencias de la Educación en general).
- «Exceptional child education resources» (todos los aspectos de la educación de niños deficientes).
- «FRANCIS» (entre otras, Ciencias de la Educación en general).
- «NICSEM/NIMIS» (materiales para niños minusválidos).
- «Special Education Materials» (material impreso y audiovisual relativo a educación especial).

De este enumeración dos puntos me parecen especialmente significativos: en primer lugar que si exceptuamos la base FRANCIS todas son americanas, y en segundo lugar, de las seis citadas, cuatro se centran en temas relativos o cercanos a la educación especial.

En cualquier caso, y al margen de las conclusiones que cada uno quiera extraer, lo que parece a todas luces evidente es que tanto por su procedencia como por su temática, de una parte van a resultar insuficientes pues hay muchas cuestiones o áreas de trabajo que no van a aparecer, o bien parte de la información de que disponen no nos va a ser de utilidad por estar centrada en los modelos y pautas del ámbito cultural donde se produce (2).

Tomando en cuenta estas consideraciones y en función del volumen de publicaciones educativas que se editan en España o en castellano, parece cada vez más imprescindible poder disponer de bases de datos propios que pongan a disposición de todos la literatura científica que se produce, no sea que lleguemos a conocer antes y mejor las realizaciones americanas o del resto de Europa que las de nuestros más inmediatos vecinos.

Las realizaciones a este nivel son escasas: el Ministerio de Educación y Ciencia a través de la Secretaría General Técnica tiene disponibles las siguientes bases distribuidas por el centro de proceso de datos y accesibles mediante la RUE:

Se trata de las bases LEDA sobre legislación educativa, BIDE sobre cursos y becas de todo el mundo para estudiantes iberoamericanos, ISOC sobre Ciencias Sociales y Humanidades en general y REDINET actualmente en fase de implantación y a la que luego me referiré.

El área de disciplinas cubiertas es bastante restringido, sobre todo por el número de referencias almacenadas, pero más preocupante es la situación referida a los tesauros. Los existentes de los que tenemos noticia son:

- de tipo general; EUDISED, Ibe Education Thesaurus, Thesaurus Pädagogik, ERIC, Thesaurus of British Educational Terms, FRANCIS y Tesoro Colombiano de la educación.
- específicos; Tesoro de formación profesional, Evaluation Thesaurus, Thesaurus: enfance inadaptée, Special Education Theaurus, Terminology of adult education, Thesaurus education permanente.

De todo este conjunto los elaborados originalmente en lengua castellana son el colombiano y el de formación profesional de Uruguay.

Además de estos, los elaborados por organismos internacionales suelen ser tesauros multilingües, siendo normalmente el castellano uno de los idiomas de los mismos. Tal es el caso del EUDISED elaborado por el Consejo de Europa o el IBE por la UNESCO.

Pero en cualquier caso hay que tener presente que se trata usualmente de traducciones y no de tesauros específicamente adaptados a la realidad educativa y cultural de cada país. Son valiosos instrumentos para asegurar la comunicación internacional, pero en su propia concepción estriba también su principal defecto. A este respecto son clarificadoras las valoraciones realizadas sobre el de

la UNESCO por diferentes usuarios de los países latinoamericanos con sistemas educativos y terminologías muy variadas. (3).

Para suplir esta carencia son varias las iniciativas tomadas al respecto pero, que sepamos, hasta la fecha ninguna se ha plasmado en realizaciones concretas.

Como antes apuntamos el CSIC a través del ISOC tiene definido un plan de trabajo para elaborar tesauros castellanos de las diversas disciplinas de su competencia. Actualmente están elaborando uno de Urbanismo, y según nos han manifestado preveen comenzar con el de educación no antes de dos años.

Junto a ello, autores como Rodríguez Delgado (4), se han manifestado en repetidas ocasiones partidarios de establecer planes conjuntos para la elaboración coordinada de tesauros en portugués y castellano.

Ya en la primera reunión de directores de centros de documentación educativa del área andina celebrada en Lima en 1978, se presentó el proyecto del «Tesauro educacional CENDIE» (Centro Nacional de Documentación e Información Educativa) con el patrocinio de la Organización de Estados Americanos.

Se realizó un estudio comparativo de los tesauros de la UNESCO, ERIC y EUDISED y se elaboraron una serie de descriptores a título experimental. Pero que sepamos, el proyecto no ha pasado de ahí.

Por último, en la sección de Pedagogía de la Facultad de Filosofía y Ciencias de la Educación de San Sebastián hemos comenzado un proyecto piloto centrado en la elaboración de un micro-tesauro referido al área de la investigación educativa, técnicas y metodología. Micro-tesauro que sería perfectamente integrable en un tesauro general de las Ciencias de la Educación en cuya elaboración participa el ICE de la Universidad del País Vasco y el centro de documentación del ITE de Alcalá de Henares, habiendo entablado conversaciones con el ISOC de cara a coordinar esfuerzos.

## **EL ÁREA DE LA INVESTIGACIÓN EDUCATIVA. TÉCNICAS Y METODOLOGÍA.**

Las razones de elegir este campo para la elaboración del primer micro-tesauro son varias; entre ellas –todo hay que decirlo– porque soy profesor del departamento de Pedagogía Experimental y es este el campo de las ciencias de la educación que menos desconozco. Pero al margen de ello, es preciso hacer notar que los que trabajamos en este campo, a la hora de realizar una búsqueda documental normalmente la realizamos porque nos interesa disponer de la información referente a lo trabajado en un tema determinado; pero no es menos frecuente que en lugar del contenido temático nos interese más disponer de las referencias de investigaciones que han desarrollado tal o cual método o diseño, éste o aquél procedimiento estadístico que nos interesa ver como opera independientemente de que sea una investigación sobre el aprendizaje lector o sobre la motivación de los alumnos.

Y para recuperar la información desde este punto de vista, los tesauros ac-

tualmente disponibles aportan un conjunto de descriptores muy restringido y poco adecuado.

Centrándonos en el EUDISED y en el de la UNESCO por ser éstos los más usuales en castellano, un cuadro comparativo sería más o menos el siguiente:

En el EUDISED de los 42 «conjuntos semánticos» que en que se halla dividido, el 16 corresponde a «Investigación». En total contiene 68 descriptores agrupados en cinco subconjuntos:

- *Métodos de investigación*, 21 descriptores como: «análisis comparativo», «método histórico», «análisis de items»...
- *Técnicas de investigación*, 19, tales como: «medición», «experimentación», «observación», «trabajo de campo»...
- *Tipos de investigación*, 16 como: «investigación aplicada», «investigación comparativa», «inv. empírica»...
- *Organización de la investigación*, 6 descriptores, entre ellos: «centro de investigación», «proyecto piloto»...
- *Método estadístico*, 6, a saber: «correlación», «covarianza», «factorial», «multivariado», «regresión» y «varianza».

Es de notar que de todos los métodos de investigación es el único que forma un grupo aparte.

El de la OIE-UNESCO considera el apartado investigación dentro del campo segundo con un total de 67 descriptores. Igualmente se subdivide en cinco:

- *Campos de investigación*, 10: «investigación curricular», «investigación del comportamiento», «pedagogía experimental»
- *Problemas*, 9: «prevención de la delincuencia», «problemas curriculares», «problema familiar».
- *Tipos de investigación*, 23: «análisis comparativo», «estudio de casos», «estudio estadístico»...
- *Metodología de investigación*, 15: «base de datos», «grupo de control», «validez»...
- *Datos*, 10: «análisis de datos», «análisis estadístico», «procesamiento de datos»...

Como puede verse si nos centramos en lo referido a la estadística, usando el tesoro EUDISED, sólo podemos distinguir entre cinco tipos de análisis, y en el de la UNESCO nos tendremos que contentar con el único término de «análisis estadístico» pues éste ni siquiera se subdivide.

En honor a la verdad hay que decir que no todos los tesauros son tan limitados como éstos, así la información aportada por ERIC es más amplia, pero repito que me he centrado en aquellos de los que existe versión castellana.

Desde el punto de vista del usuario esto supone que las búsquedas que se realicen van a ser de muy baja precisión, y que para obtener la información de-

seada vamos a vernos en la necesidad de extraer casi todas las referencias existentes agrupadas bajo dichos epígrafes. Cualquier resultado va a llevar implícito mucho «ruido», pues la mayoría de las referencias no van a ser pertinentes, con lo que ello lleva aparejado de costo adicional e ineficacia del sistema.

Evidentemente, la solución a esto es elaborar un tesoro, o mejor dicho, un micro-tesoro específicamente diseñado para cubrir y fijar la terminología de este ámbito con el nivel de precisión y exhaustividad necesario para solventar las necesidades planteadas.

Como vemos, centrados en el castellano y en las Ciencias de la Educación, de una parte las bases de datos existentes son pocas y cubren un espectro temático limitado, y de otra los tesauros existentes son traducciones de los multilingües resultando en ocasiones inadecuados.

Como un intento de contribuir a ir remediando esta situación, se inscribe el proyecto REDINET del que seguidamente paso a describir sus características más importantes.

## **EL PROYECTO REDINET. CARACTERÍSTICAS GENERALES.**

Teniendo en cuenta las cuestiones antes formuladas y considerando además el continuo incremento del número de investigaciones sobre educación que en España se realizan, el CIDE (organismo impulsor del proyecto) consideró que era necesario instrumentar un sistema de información que asegurase la mutua comunicación entre todos los investigadores. Necesidad que se ve más justificada al considerar la diversidad de instituciones que producen o financian trabajos de este tipo (Facultades y Secciones de Pedagogía, Escuelas de formación del Profesorado, ICEs, Consejerías de Educación de las Comunidades Autónomas...

La primera decisión fue la referida a estructurar la organización según el modelo de las Comunidades Autónomas, de tal manera que más que de una base de datos, hay que hablar de una red de bases de datos integrada por el CIDE y las 17 Consejerías de Educación o instituciones en que éstas hayan delegado.

Se considera que cada miembro de la Red es autónomo en su demarcación, pudiendo, si su infraestructura y medios lo permiten, formar a su vez su propio servicio de documentación.

Para asegurar el principio de reciprocidad e igualdad entre los miembros, cada uno se compromete a remitir al Centro de Proceso de Datos del Ministerio la información documental referida a las investigaciones producidas en su zona con el fin de que esté disponible para el resto.

Se podría hablar pues de un servicio de documentación descentralizado, o mejor semi-centralizado, en la medida en que los integrantes del sistema no se limitan a priori a ser únicamente terminales del ordenador central, aunque por obvias razones económicas es éste el que almacena la totalidad de la documentación generada por cada miembro poniéndola a disposición de toda la Red.

De esta manera cada miembro adecúa su modo de operar a los fondos de que dispone y a las necesidades y demandas que se le planteen, de forma que habrá centros a los que les compensará más operar como terminales del ordenador central en la mayoría de los casos, mientras que en otros el funcionamiento será normalmente autónomo.

Como contrapartida esto exige que la infraestructura informática sea compatible y sobre todo disponer de formatos y lenguajes documentales comunes a todos para asegurar las intercomunicación, máxime si consideramos que si bien el ordenador central dispone de todas las referencias enviadas por cada miembro, las fuentes primarias se encuentran solamente allá donde se han producido, de tal manera que una búsqueda documental debe proporcionar además de la información habitual, la necesaria para que el usuario pueda localizar los documentos que le interesen.

Centrándonos ahora en lo referido a los formatos, la ficha para realizar el vaciado e «indización» documental consta de dos partes:

La primera contiene los datos de información general para identificar y localizar el documento original (título, autor, datos editoriales si está publicado o datos del centro donde es localizable en caso contrario etc).

Los campos de la segunda parte se basan en lo ya comentado. Tan útil es la información temática de un trabajo como la referida a las técnicas y metodología del mismo. Así se incluyen campos como los objetivos de la investigación, la muestra o el objeto del trabajo, los instrumentos de obtención de información, las técnicas de análisis etc, concluyendo con los descriptores temáticos y metodológicos.

Esta ficha es el primer elemento de la necesaria normalización de la Red. Por lo que se refiere al segundo, la comisión asesora optó por utilizar en una primera fase el tesoro EUDISED pues pensamos que su terminología y su sistema de clasificación de las familias semánticas resultaban más adecuados.

En cualquier caso ya hemos comentado que presenta problemas, y entre ellos el referido a las cuestiones metodológicas. Para remediar en lo posible esta cuestión actualmente se dispone de listas controladas de términos que provisionalmente pueden ser útiles hasta que no se disponga del o de los tesauros necesarios.

Por último señalar que el logical que soporta el sistema documental es el paquete UNIDAS desarrollado por la casa UNIVAC. Este permite recuperar documentos a través de tres áreas:

- a) aspectos significativos del documento (lo que hemos llamado información de tipo general: autor, tipo de investigación, año).
- b) descriptores (en nuestro caso de ambos tipos).
- c) búsqueda en texto libre (todos los campos destinados a la descripción pormenorizada de las características metodológicas).



En cualquiera de las tres formas de recuperación de la información o en sus posibles combinaciones se acota el conjunto documental mediante los operadores booleanos habituales.

## CONCLUSIONES FINALES

Referido a la documentación científica automatizada en el campo de las Ciencias de la Educación se observa el bajo número de bases de datos existentes con un campo temático limitado y la inexistencia de tesauros o micro-tesauros elaborados originalmente en lengua castellana y adecuados al marco socio-cultural español.

Esto conlleva una fuerte dependencia externa y un bajo nivel de comunicación entre los miembros de la comunidad científica. (Como nota complementaria digamos que nos ha sido imposible disponer de los tesauros colombiano o uruguayo y en general tener conocimiento preciso de las realizaciones y proyectos planteados en Latinoamérica).

La primera línea de acción para solventar estas carencias pasa por aumentar el nivel de comunicación entre los investigadores con el fin de lograr un acuerdo y poder disponer de un lenguaje documental común por mínimo que éste sea.

El estado actual de la tecnología informática y en general de la investigación sobre tratamiento automático de la información implica el uso de lenguajes controlados y normalizados, y mientras esto no cambie lo anterior es un requisito previo para cualquier acción ulterior.

Logrado esto sería posible generar pequeñas bases de datos locales o sectoriales fácilmente interconectables entre sí a un costo relativamente bajo, pues el desarrollo de la micro-informática lo posibilita.

## NOTAS

1. FUINCA, (1981) *Bases de Datos del Mundo*. Alhambra. Madrid.
2. VILLAGRÁ, A., y ROMÁN, A., (1981). «Diferente utilidad de las bases de datos americanas y europeas en las ciencias sociales» *Rev. Esp. Doc. Cient.* Vol. 4, Número 2, pp. 112-129.
3. Cfr. Los siguientes artículos publicados en la *Revista de la UNESCO de Ciencias de la Información, Bibliotecología y Archivística*:

BRAHM, L., y GUTIÉRREZ, G., (1983): «REDUC: una red de documentación en educación al servicio del desarrollo en América Latina». Vol. V, Número 2, Abril-Junio. pp. 93-97.

FURIADO, A., (1983) «La Red Internacional de Información Educacional (INED) y la participación de America Latina y el Caribe». *Ibid.* pp. 86-89.

4. RODRÍGUEZ DELGADO, R., (1979): «Plan para la elaboración de Tesoros coordinados en portugués y español». *Rev. Esp. Doc. Cient.* Vo. 2, n.º 4, pp. 337-355

RODRÍGUEZ DELGADO, R., (1980): «La integración de las lenguajes documentarios, fin de Babel». *Rev. Esp. Doc. Cient.*, Vol. 4, n.º 3, pp. 333-341.



## librería herder

SERVICIO INTERNACIONAL

les ofrece su sección de:

- **Psicología**
- **Pedagogía**
- **Psicoanálisis**
- **Tests**

Con unos 5.000 títulos en stock permanente

Solicite el envío gratuito del boletín bibliográfico de estos temas o, mejor, visite nuestra exposición.

**LIBRERIA HERDER**  
Balmes, 26 - Tel. 317 05 78  
08007 BARCELONA