

UNIFICACION DE LENGUAJES DOCUMENTARIOS. BASES PARA LA CREACION DE TESAuros COMUNES EN ESPAÑOL *

INTRODUCCIÓN

Los lenguajes bibliotecológicos, cuya difusión general comenzó hace más de cien años, constituyeron un importantísimo instrumento para ordenar conceptualmente los fondos monotemáticos de las bibliotecas y facilitar la recuperación de los libros al servicio del lector, pero resultan hoy insuficientes y han dado lugar a la proliferación de tesauros que son utilizados en sectores especializados del conocimiento.

Sin embargo, por estar concebidos los tesauros para satisfacer necesidades particulares de organismos internacionales especializados, de industrias determinadas, de asociaciones profesionales, o incluso de pequeños centros sectoriales de documentación, sus procesos de elaboración, su estructura y sus campos temáticos se han creado en mutuo aislamiento y en peligrosa incomunicación.

Existe, además, una total ausencia de normalización de los tesauros, que dificulta o hace imposible el acceso de diversas fuentes de información desde cualquier punto central, ya que es necesario utilizar instrumentos y codificaciones muy distintos para buscar el mismo tipo de información, o para pasar de un sistema a otros sistemas.

La necesidad de establecer a escala internacional lenguajes y sistemas documentarios comunes es ampliamente reconocida por la UNESCO, por la Junta Interorgánica de las Naciones Unidas (IOB) y por la Federación Internacional de Documentación (FID).

Pero para ello no solamente hay que llevar a cabo una división más lógica de los distintos ámbitos del conocimiento, sino que también es necesaria una codificación numérica normalizada. La creación de un sistema general de codificación permitiría el tratamiento uniforme, tanto manual como mecanizado, de la información contenida en cualquier tipo de documento, facilitando extraordinariamente su recuperación.

Para elaborar el propuesto instrumento conviene explorar las posibilidades metodológicas de la teoría de sistemas generales, cuya actual expan-

* Extracto de la ponencia presentada por el autor a REUNIBER - 78.

sión a todos los campos del saber y a toda clase de aplicaciones ofrece un camino realmente prometedor.

Proyecto de tesoro común

Partimos de la hipótesis de elaborar un instrumento de estructura decimal, con seis niveles, más listas edicionales, codificado numéricamente para su tratamiento mecanizado y la fácil traducción automática de los descriptores, intentando la síntesis de los principios de organización de la CDD y la CDU, por una parte, y de los tesauros, por otra.

El primer nivel, de un dígito, podría tener la siguiente estructura:

- 0.—Generalidades.
- 1.—Cosmología. Astronomía. Ciencias de la Tierra.
- 2.—Ciencias físicas.
- 3.—Ciencias químicas.
- 4.—Biología. Psicología. Ecología.
- 5.—Ciencias humanas.
- 6.—Tecnología (ciencias aplicadas).
- 7.—Literatura. Arte.
- 8.—Concepciones del Universo y del Hombre. Religión. Filosofía.
- 9.—Marco espacio-temporal. Geografía. Historia.

El segundo nivel de la clase 5. Ciencias humanas, por ejemplo, podría tener esta forma:

- 50.—Documentación en ciencias humanas. Estadística social.
- 51.—Educación. Cultura. Civilización.
- 52.—Ciencias políticas.
- 53.—Ciencias económicas.
- 54.—Derecho. Ciencias jurídicas.
- 55.—Administración pública.
- 56.—Trabajo. Asistencia pública.
- 57.—Ciencias militares.
- 58.—Sociología. Población. Demografía.
- 59.—Etnología. Folklore.

El tercer nivel del grupo "54. Derecho. Ciencias jurídicas", podría estar así configurado:

- 540.—Documentación en ciencias jurídicas. Fuentes.
- 541.—Derecho en general.

- 542.—Derecho Internacional.
 543.—Derecho Público. Derecho Constitucional.
 544.—Derecho del Trabajo.
 545.—Derecho Penal.

 549.—Historia del Derecho. Geografía jurídica.

Como se observa, una codificación normalizada puede conectarse fácilmente con cualquier sistema. En el caso del Derecho, la clasificación de la CDU es bastante lógica —lo que no sucede en otros campos—, por lo que podría ser utilizada en gran medida por un sistema común.

En las ciencias naturales, evidentemente, habría que conceder primacía a los tesauros ya desarrollados y de amplio uso, los que podrían conectarse con el sistema común al nivel del tercer dígito, sin pérdida de su estructura.

Sin embargo, en aquellos campos en los que no hubiera tesauros adecuados, podría emplearse el sistema común hasta el sexto dígito, que representa una capacidad teórica de un millón de términos.

La codificación numérica común podría aplicarse fácilmente a los tesauros que carecen de ese tipo de codificación, los que constituyen la mayoría de los existentes.

El sistema común se completaría con “listas” de descriptores o identificadores de igual nivel.

Por ejemplo, bajo el descriptor “Alcantarillado”, cuyo número de código podría ser “611.446”, iría una lista de la siguiente estructura:

Alcantarillado combinado	611.446-003
Alcantarillado de aguas de lluvia	611.446-002
Alcantarillado sanitario	611.446-001
Etcétera.	

En estas listas el número de orden puede ser significativo, por ejemplo, si se ordenan los “elementos químicos” por peso atómico, o no serlo, lo que da una gran flexibilidad al sistema.

Bases para la creación de tesauros comunes en castellano

En 1970 la UNESCO recibió un informe elaborado por un destacado grupo de expertos, en el que se contenían veintidós recomendaciones destinadas a poner en práctica el programa UNISIST. La Recomendación 4 —que juntamente con otras se consideró de máxima prioridad— se refiere

al difícil sector de la especificación temática en los sistemas de información científica, que no supone únicamente problemas de carácter epistemológico y práctico relativos a la clasificación de las ciencias, sino también la normalización terminológica entre los países y entre los múltiples campos científicos, a fin de crear vocabularios controlados o tesauros destinados a los sistemas de información y efectuar progresos en la indización automática de los documentos. También se subraya la necesidad de llevar a cabo un esfuerzo concentrado para explorar la viabilidad de integrar los tesauros de los diversos sectores científicos¹.

En el ámbito del castellano, paralelamente, se comenzaron a llevar a cabo, en el año 1970, trabajos en el seno del Centro Latinoamericano de Documentación Económica y Social (CLADES) de la Comisión Económica de las Naciones Unidas para América Latina (CEPAL). El CLADES, tras traducir el Macrotesauro de la Organización de Cooperación y Desarrollo Económicos (OCDE)², elaboró un sencillo sistema de recuperación de información y de indización bibliográfica mediante computadora³, obteniéndose como primer resultado una bibliografía analítica experimental mecanizada⁴. Ya entonces se indicó la conveniencia de aplicar la teoría de sistemas a la elaboración de tesauros.

En 1973 se llevó a cabo otro proyecto en el Instituto de Cultura Hispánica, en Madrid, cuya Unidad de Informática Documentaria utilizó las experiencias obtenidas en el CLADES y publicó en 1976 un Listado de Descriptores elaborado en computadora⁵, en colaboración con el Centro de Documentación Iberoamericana y con un grupo de expertos en informática. En este trabajo se efectuaba una comparación estadística entre los descriptores procedentes del Macrotesauro de la OCDE, los adicionales obtenidos por el Centro de Documentación y los términos correspondientes de la CDU, poniéndose de relieve los problemas hallados y la necesidad de crear un tesoro iberoamericano de amplio alcance⁶.

¹ En relación con el proyecto UNISIST se está estudiando, por grupos de trabajo FID/SRC, un nuevo sistema de ordenación (*Broad System of Ordering*).

² CEPAL, CLADES. *Macrotesauro OCDE*. Versión preliminar. Vols. I y II. Santiago de Chile, 1973.

³ RAFAEL RODRÍGUEZ DE CORA y JUAN MORANDÉ O'REILLY: *La computadora y sus sistemas y programas aplicados a la documentación* (CLADES/WG. 1/L. 6), *Descriptores del 14.º período de sesiones de la CEPAL procesados en computadora* (CLADES/WG. 1/L. 18), Santiago de Chile, 1971.

Manual para usuarios del KWIC/360. Sistema de recuperación e indización bibliográfica (CLADES/INS/2), CEPAL, Santiago de Chile, 31 de marzo de 1972.

⁴ CEPAL, CLADES, *Bibliografía analítica del Instituto Latinoamericano de Planificación Económica y Social (ILPES)* (CLADES/BBG/4), Santiago de Chile, 24 de marzo de 1972.

⁵ Instituto de Cultura Hispánica. *Listado de Descriptores Económicos, Sociales y Políticos - 1976*, Introducción por Rafael Rodríguez Delgado, Instituto de Cultura Hispánica, Madrid, 1976.

⁶ Se observó una gran falta de correspondencia entre la estructura de la CDU y la

En 1977 el Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente (CEPIS) convocó en Lima, con la colaboración del Centro Internacional de Investigaciones para el Desarrollo del Canadá, a un Grupo Consultivo⁷, al cual se presentó un Documento de trabajo en el que se aplicaba por primera vez la teoría de sistemas a la clasificación documentaria y a la elaboración de tesauros, obteniéndose un modelo de tesoro por medio de minicomputadora en un plazo de dos meses⁸.

A principios de 1978, por último, el Consejo Nacional de Investigación del Perú, con la cooperación de la Organización de los Estados Americanos (OEA), organizó un simposio destinado a establecer las líneas generales de una Red Nacional de Información Científica y Tecnológica (RENICYT), al que se presentó un documento en el que también se aplicaba la teoría de sistemas a la construcción de tesauros⁹. Uno de los resultados del simposio fue el de constituir un grupo interdisciplinario encargado de estudiar las posibilidades de correlación entre los tesauros elaborados o en proyecto en el Perú y de examinar la posibilidad de crear un megatesoro, o tesoro común¹⁰.

Por otra parte, el estudio de un lenguaje científico en castellano destinado a los sistemas de recuperación de información constituye una tarea básica de investigación para el Centro Nacional de Información y Documentación Científica (CENIDOC), creado en España en 1975 en el seno del Consejo Superior de Investigaciones Científicas (CSIC)¹¹.

Los diversos y valiosos trabajos de investigación en informática documentaria llevados a cabo por la Fundación para el Desarrollo de la Función Social de las Comunicaciones (FUNDESCO), en España, y la creación por ésta de una red automatizada para la información científica y técnica (Red INCA)¹², constituyen importantes bases teóricas y prácticas para que un

empírica del Macrotésoro OCDE. El Centro de Documentación del ICH obtuvo más de un 50 por 100 de nuevos descriptores, que no se hallaban en el Macrotésoro.

⁷ CEPIS - CIID. Grupo consultivo sobre Información en Ingeniería Sanitaria y Ciencias Ambientales, Lima, 14-16 noviembre 1977.

⁸ CEPIS - CIID. *Bases sugeridas para la creación de un Microtesoro de términos sobre Ingeniería Sanitaria y Ciencias del Ambiente (MISCA)*. Documento de trabajo núm. 2 del Grupo Consultivo sobre Información en Ingeniería Sanitaria y Ciencias Ambientales. Lima, 14-16 noviembre 1977.

⁹ RAFAEL RODRÍGUEZ DELGADO: *Bases para la utilización de lenguajes documentarios en la RENICYT*. Documento de trabajo núm. 4, Simposio RENICYT - Red Nacional de Información Científica y Tecnológica - Consejo Nacional de Investigación, Programa CNI-OEA. Lima, marzo 1978.

¹⁰ Bajo los auspicios del proyecto RENICYT se constituyó en Lima un grupo encargado de estudiar la normalización de la elaboración de tesauros, especialmente en el campo de la ciencia y la tecnología.

¹¹ Centro Nacional de Información y Documentación Científica. Consejo Superior de Investigaciones Científicas. Folleto informativo. Valencia, 1977.

Una de las cuatro áreas en que se centran sus programas es el "Estudio del lenguaje científico en castellano para los sistemas de recuperación de la información". (Pág. 3).

¹² Red de Información Científica Automatizada INCA. Información básica. Madrid, s. a. 1978?

lenguaje documentario común pueda tener inmediata aplicación a escala internacional, tanto dentro del mundo de habla castellana como en cuanto a la intercomunicación con los grandes sistemas que utilizan como vehículo otros idiomas.

Estos antecedentes, que constituyen sólo una muestra de la experiencia existente en elaboración de tesauros en español y en recuperación mecanizada de la información en el ámbito iberoamericano, permiten afirmar la viabilidad de la tarea de crear tesauros comunes en castellano. Un esquema para establecer un plan modular de acción orientado hacia ese fin podría ser el siguiente:

Objetivos generales

1. Crear un lenguaje documentario común en castellano, aplicable a la elaboración de tesauros generales y especializados y a la recuperación e intercambio de información en redes mecanizadas.
2. Establecer, mediante códigos numéricos uniformes, compatibilidad entre los tesauros existentes o de nueva creación para lograr su fácil intercomunicación.
3. Crear un sistema de traducción automática de descriptores codificado numéricamente de modo que se facilite la intercomunicación entre tesauros en distintos idiomas.

Plan de acción preliminar

Un plan de acción sencillo y viable sería el siguiente:

- Acuerdo de las organizaciones interesadas en el proyecto para su ejecución en el plazo más breve posible.
- Nombramiento de un Comité Asesor del proyecto cuya primera misión sería la de establecer un plan de trabajo a corto plazo.
- Aprobación por los organismos participantes del plan de trabajo, que podría consistir en:
 - a) Investigar las opciones existentes para crear un lenguaje documentario común en castellano.
 - b) Creación de tesauros, a fin de poner a prueba las hipótesis que se establezcan. Bazaría con elaborar tres tesauros de ensayo, con

- menos de mil términos cada uno, a fin de hacer viable la realización del proyecto a corto plazo.
- c) Establecer un sistema de codificación numérica integrada, aplicable a los tesauros de ensayo.
 - d) Indizar materiales documentarios de prueba con esos tesauros.
 - e) Llevar a cabo pruebas de recuperación e intercambio de información entre los sistemas conceptuales representados por los indicados tesauros.
 - f) Traducción al inglés, francés y portugués de muestras de los descriptores obtenidos en castellano, y recuperación mecanizada de información en esos idiomas con ayuda de los respectivos tesauros multilingües.

Dada la experiencia acumulada existente —que supone una cuantiosa inversión— y las infraestructuras disponibles, se estima que el proyecto podría llevarse a cabo en año y medio contando con la cooperación en personal, equipo y logical de los organismos participantes. Convendría contar con un coordinador, asistido por una secretaria, y con un local suministrado por alguno de los organismos nacionales o internacionales participantes en el que se centralizase el funcionamiento del proyecto.

Los resultados obtenidos serían objeto de evaluación por los organismos participantes, que decidirían ulteriormente acerca de la aplicación en gran escala del sistema, y de su difusión en los países de habla castellana.

RAFAEL RODRIGUEZ DELGADO