

TESAURO DE EDUCACIÓN SUPERIOR: Un ensayo de experiencia interdisciplinar

Por : Javier Gimeno Perelló

Correo electrónico: jgimeno@rect.ucm.es

Lingüista y bibliotecario. Universidad Complutense de Madrid – España.

Lima, febrero de 2002

Resumen

Se realiza una reflexión de carácter lingüístico y filosófico sobre el papel desempeñado por los lenguajes documentales asociativos en la organización del conocimiento científico y en su recuperación y difusión. Se subraya la evolución de estos lenguajes en herramientas lingüistas de carácter transdisciplinar por cuanto las nuevas tecnologías de la información exigen instrumentos capaces de imbricar diferentes especialidades científicas en el ámbito de la sociedad del conocimiento. En este contexto, se ha elaborado el Tesauro de Educación Superior, del cual se describe su metodología, características, fuentes documentales, composición y estructura.

Introducción. De los lenguajes documentales asociativos a la transdisciplinariedad.

Una dificultad que se presenta a la hora de buscar información sobre una determinada disciplina es la selección de la terminología apropiada que facilite el acceso directo a la información y la adecuada combinación de los términos que permita una búsqueda precisa. Ambas cuestiones se pretenden solventar con los lenguajes documentales. A diferencia de las nomenclaturas tradicionales, los léxicos elaborados con vistas a la recuperación de la información de un campo del conocimiento, como los tesauros, estructuran el vocabulario científico en tres direcciones: sinonimia, jerárquica y asociativa, de suerte que cada término dispone de sus correspondientes relaciones semánticas. En este sentido, un lenguaje documental cumple dos cometidos esenciales: uno, de carácter teórico, como herramienta lingüística de estructuración lexicográfica del campo semántico; otro, de carácter práctico,

como herramienta documental de indización o selección de palabras-clave o descriptores, y de búsqueda, acceso y recuperación de la información en ficheros automatizados y bases de datos documentales, sean estos residentes -accesibles fuera de línea, tales como bases de datos almacenadas en disco duro del propio ordenador, en cd-rom o disquete-, o telemáticas - accesibles en línea, mediante conexión a redes locales, nacionales, etc., Internet o Intranet-.

Los nuevos sistemas de clasificación representados en los lenguajes documentales no jerárquicos de estructura asociativa reflejan mejor que las tradicionales clasificaciones enciclopédicas y jerárquica la transdisciplinariedad del conocimiento científico. Reflejan asimismo la interdisciplinariedad que las nuevas tecnologías de la información y de las comunicaciones facilitan en la nueva estructura informacional de la ciencia, aún a pesar de la fuerte especialización que se experimenta en los centros educativos y de investigación.

Los nuevos lenguajes de clasificación documental contruidos sobre el concepto de categoría como principio de representación y organización del conocimiento, basados en la asociación de ideas y de relación conceptual, alejados de la jerarquización y de la fragmentación, permiten una estructuración multidimensional y transdisciplinar del conocimiento. Sistemas soportados en redes neuronales (Neurosoc), algoritmos genéticos (CGO o COBLI), modelos conceptuales de recuperación de objetos multimedia, aplicaciones de lógica terminológica de búsqueda documental (Web/TwebS), junto con nuevas disciplinas técnico-científicas encaminadas a la introspección de las regiones internas de cada disciplina –“intra geografía” (Esteban Navarro), o al estudio de las fronteras entre las ciencias –“kinegeografía” (ibidem)- , etc., constituyen un nuevo paradigma epistemológico de transdisciplinariedad superador de la fractura y fragmentación del conocimiento, en la búsqueda de la integración de las ciencias físico-naturales y antro-po-sociales, en una cosmovisión bio-physis-anthropos-social (Edgar Morin).

Hempel se refería en 1966 al concepto multidimensional de la ciencia: “la sistematización científica requiere el establecimiento de diversas conexiones, mediante leyes o principios teóricos, entre diferentes aspectos del mundo empírico, que se caracterizan mediante conceptos científicos. Así, los conceptos de la ciencia son nudos en una red de interrelaciones sistemáticas en la que las leyes y los principios teóricos constituyen los hilos. [...] Cuantos más hilos converjan o partan de un nudo conceptual, tanto más importante será su papel sistematizador o su alcance sistemático”¹.

Relevantes fenómenos como el hipertexto, los hiperdocumentos, los documentos electrónicos, interactivos en soporte digital y en red, la biblioteca virtual, etc., productos todos ellos de las NTIC, rompen la concepción espacial de los tradicionales soportes para adentrarse en la temporalidad y espacialidad de los soportes nuevos. Las NTIC han permitido la liberación de toda la potencia y capacidad del hipertexto y todos los demás soportes electrónicos de información.

¹ Hempel, C.G.: Philosophy of natural science. Prentice-Hall, 1966. Cit. por Izquierdo Arroyo, op.cit. p.350

El hipertexto rompe la linealidad del lenguaje (Jakobson) y lo convierte en no secuencial en virtud de lo que T.H. Nelson denomina *almacenamiento xanalógico*², definido como la interrelación de partes de documentos primarios para constituir nuevos documentos, permitiendo que “las distintas unidades se construyan a partir de partes procedentes de otras unidades y se puedan enlazar en múltiples direcciones” (Izquierdo Arroyo, p.347).

La estructura hipertextual es equivalente a la estructura de la memoria humana, pues ambas no son lineales como el lenguaje hablado o escrito, sino multidireccionales. Tanto las ideas y los conocimientos como los elementos del hipertexto son relacionales, unos conducen a otros; nuestra mente, al igual que los enlaces hipertextuales, establecen múltiples direcciones hasta el infinito.

Los lenguajes documentales asociativos y no jerárquicos favorecen esta estructura multidimensional y multidireccional de nuestra memoria al ofrecer una estructura polijerárquica y facetada de los conceptos. Esta concepción multijerárquica, asociativa y facetada de los lenguajes documentales está presente también en el hipertexto, que permite enlazar términos, conceptos, ideas entre sí, rompiendo la unidireccionalidad lineal clásica de la estructura informativa y favoreciendo la estructura multidimensional en red .

El hipertexto trasciende el texto unitario y lo interconecta con otros textos que a su vez se interconectan con muchos otros hasta formar una gran red textual que trasciende con mucho los límites de un trabajo científico o de cualquier otro documento. A su vez, las redes textuales trascienden a la propia red para imbricarse en otras redes, en la red de redes.

“[...] los márgenes de un libro no están jamás neta ni rigurosamente cortados; más allá del título, las primeras líneas y el punto final, más allá de su configuración interna y la forma que lo autonomiza, está envuelto en un sistema de citas de otros libros, de otros textos, de otras frases, como un nudo en una red.”³

La noción de hipertexto es una noción literaria: Rayuela (Cortázar); El jardín de los senderos que se bifurcan, El Aleph (Borges), Ulysses (Joyce), “son distintas muestras de un mismo esfuerzo por escapar a la linealidad del texto tradicional”⁴. Derrida, Barthes o Foucault encontraron en el hipertexto el punto de inflexión donde reformular la relación autor-lector.

El Tesoro de Educación Superior de la Universidad Complutense de Madrid

Con vocación de transdisciplinariedad se ha elaborado el Tesoro de Educación Superior (en adelante, TES) Es un vocabulario documental que reúne cerca de dos mil términos relacionados con el mundo universitario y

² Nelson, T.H.: Interactive systems and the design of virtuality. Creative computer, v.6, nº 11, nov. 1980

³ Foucault, M.: La arqueología del saber. México: Siglo XXI, 1990, p.37

⁴ Herrero Solana, V.F.: Hiperdocumentos referenciales. Buenos Aires: Nuevo Paradigma, 1998, p. 38

científico. Si exceptuamos el Tesoro Europeo de Educación, de la Comisión Europea, el Tesoro de la UNESCO y el Tesoro Catalá d'Educació, de la Universidad Autónoma de Barcelona, no existen más léxicos científico-documentales especializados en educación y enseñanza. Y el que traemos aquí a colación, el TES de la Universidad Complutense de Madrid es el único sobre enseñanza superior y universitaria. Surge por la necesidad de estructurar el campo lexicográfico de estas áreas del conocimiento que permita una mejor sistematización del análisis documental y de la búsqueda y recuperación de la información. Es el resultado del proceso de indización documental de información relacionada con la universidad y la educación superior, llevado a cabo por el servicio de documentación del Departamento de Análisis y Planificación de la citada universidad.

Elaborado según la norma ISO 2788-1986: *Principios directrices para la elaboración y desarrollo de tesauros monolingües*, Ginebra, 1986, el TES es un lenguaje documental aplicado al análisis de la información relacionada fundamentalmente con la gestión, planificación y política educativa, con un enfoque destinado a la educación superior y universitaria en particular. Estructura, por consiguiente, el léxico utilizado en áreas temáticas de estas disciplinas como la política y gestión universitaria, sistema y administración educativa, política científica y de investigación, política cultural, política de la información y la comunicación, y demás áreas directa o indirectamente relacionadas con la educación superior y universitaria.

Para la realización y posterior mantenimiento y gestión del TES se ha utilizado el programa informático *Beat* versión 0.9 de administración de tesauros, que ha facilitado la labor de estructura e indización de descriptores.

La permanente evolución y desarrollo de la lexicografía aplicada a éstas y a cualesquiera otras disciplinas supone la continua modificación del TES en sucesivas ediciones que irán surgiendo en su doble formato impreso y electrónico, éste disponible también por Internet.

1. Fuentes del TES

Los términos que constituyen el TES proceden de cinco tipos de fuentes de información:

- a) documentos de carácter primario, tales como artículos de revistas científicas y de información general sobre enseñanza y universidad, noticias de prensa, monografías, publicaciones de universidades e instituciones dedicadas a la política y gestión educativa, universitaria y científica
- b) otros documentos primarios: literatura gris como informes y documentos de trabajo no publicados de instituciones y organismos públicos y privados de carácter nacional e internacional

- c) documentos secundarios y referenciales como guías y directorios de universidades, memorias, anuarios, planes de estudio, etc.
- d) otros documentos secundarios como bibliografías y bases de datos de educación (*BIBE: Boletín Internacional de Bibliografía sobre Educación, ERIC Database, Bases de datos del CIDE, del Consejo de Universidades, Bulletin Signalétique Sciences de l'Éducation, etc.*)
- e) lenguajes documentales, en especial el *Tesaurus Europeo de la Educación* de la Comisión de las Comunidades Europeas y Consejo de Europa, *ERIC Thesaurus, Lista de Encabezamientos de Materia* de la Biblioteca de la Universidad Complutense de Madrid, *Tesaurus de la UNESCO, Tesaurus de las Comunidades Europeas, etc.*

Para la elaboración de este tesaurus se ha empleado una metodología mixta de pre y postcoordinación lingüística. Seleccionando la terminología procedente de la documentación indizada para las bases de datos del servicio de documentación del DAP UCM, se ha contrastado aquella con lenguajes documentales y publicaciones universitarias de carácter secundario, añadiendo al TES aquellos descriptores imprescindibles para su estructuración lingüística.

2. Composición y estructura

La terminología del TES se compone de descriptores e identificadores. Los descriptores, que conforman el núcleo principal del tesaurus, se estructuran de manera jerárquica, alfabética y asociativa. El TES consta, en su conjunto, de cuatro partes diferenciadas:

Primera parte. Microtesauros o facetas bajo las cuales se estructuran los descriptores según áreas temáticas, que pretenden reflejar la interdisciplinariedad del tesaurus, divididas éstas en siete facetas subdivididas a su vez en tres niveles o subfacetas:

- Sistema y política educativa, que incluye a su vez un nivel relevante por cuanto reúne los términos relativos a la educación superior y universitaria procedentes de la indización de documentos de estas especialidades y de tesauros y lenguajes como el Tesaurus de Educación de la Comisión de las CC.EE., el Tesaurus ERIC o la *Lista de Encabezamientos de Materia* de la Biblioteca de la UCM
- Ciencia, investigación y cultura, cuyos descriptores proceden de documentos de destacable implicación en el ámbito universitario al referirse, en su mayoría, a los distintos campos del conocimiento científico, tecnológico y cultural, con una mención expresa a la terminología de la investigación, y lenguajes documentales al uso, en especial el *Tesaurus de la*

UNESCO, la Lista de Encabezamientos de Materia de la Biblioteca de la UCM, o la CDU

- Información y comunicación, creada a partir de descriptores extraídos de documentos cuyo contenido en este campo es también relevante para el ámbito principal del tesoro, y del *Tesoro de la UNESCO* y otros vocabularios científicos
- Política y Administración Pública, ámbitos igualmente relacionados con la universidad y la enseñanza pública cuya terminología se ha confeccionado a partir de la documentación indizada y de lexicografía documental al uso (*Tesoro de las Comunidades Europeas* y otros)
- Aspectos económicos y legales, de especial relevancia en el campo educativo y científico, cuyos descriptores proceden de la indización elaborada y de lenguajes como el *Tesoro de la OCDE*, *Tesoro de la OIT* y términos jurídicos extraídos de vocabularios especializados en terminología legislativa y jurídica y boletines oficiales
- Medio ambiente y recursos físicos con la terminología extraída de aquellos documentos indizados más directamente vinculados a la temática específica del tesoro y contrastados con otros descriptores, en especial de las bases de datos del CINDOC
- Sociedad, faceta pluridisciplinar cuya estructuración responde a documentación incluida en las bases de datos de contenido anejo o relacionado con el mundo educativo, científico y universitario y para cuya estructuración se han utilizado también fuentes secundarias como la CDU, el *Tesoro de las Comunidades Europeas*, *Tesoro de la OCDE*, *Tesoro de la UNESCO*, etc.

Para facilitar la consulta de los descriptores pertenecientes a una faceta se incluye, al final de cada una, la relación alfabética de sus correspondientes términos.

Segunda parte. Es el núcleo del tesoro y se compone de todos los descriptores y sinónimos que conforman el TES, ordenados alfabéticamente y estructurados con sus diferentes relaciones de especificidad, sinonimia, jerárquicas y asociativas, identificadas cada una de ellas por sus correspondientes operadores.

- Nota de Alcance (operador NA) que orienta, especifica o acota el uso de descriptores que pueden tener cierta ambigüedad o exigir una concreción determinada para su utilización en las búsquedas o en la indización de documentos

Evaluación educativa

NA Úsese bajo los conceptos de evaluación de profesores, docencia, programas y sistema educativo



- Sinonimia (operador USE) que remite de un término no admitido a un descriptor sinónimo

Abandono de estudios

USE Abandono escolar

A su vez, se indica mediante el operador UP (Usado Por) el término no admitido correspondiente al descriptor

Abandono escolar

UP Abandono de estudios

- Término Genérico a Amplio (operador TA) que indica el o los descriptores jerárquicamente superiores por su generalización semántica y su utilización en conceptos de carácter más amplio

Actividades escolares

TA Administración educativa

- Término Específico (operador TE) que indica el o los descriptores jerárquicamente inferiores por su concreción semántica y su utilización en conceptos de carácter más específico

Matrícula

TE Matrícula de honor

Matrícula de selectividad

- Término Relacionado (operador TR) que indica el o los descriptores que por su significado o uso mantienen una relación de carácter asociativo u horizontal e informan de otras posibilidades de indización o búsqueda de información

Plazas educativas

TR Asignación de plazas

Límite de plazas

Matrícula

Tercera parte. Índice *Kwic* o permutado: índice alfabético de todas las palabras que componen los descriptores y sinónimos del tesoro

Cuarta parte. Índice de identificadores o nombres de entidades, instituciones y organizaciones utilizados en la indización. Se incluye una lista de acrónimos, sean o no identificadores, con su desarrollo correspondiente; las siglas que a su vez son identificadores se encuentran incluidas también en el índice de éstos. Finaliza esta cuarta parte con una relación de todas las universidades españolas, públicas y privadas, y sus correspondientes siglas de identificación utilizadas por el Consejo de Universidades.

Ficha técnica

La 1ª edición del TES contiene el siguiente número de términos y relaciones:

1690 términos

1615 descriptores

2972 palabras clave (1356 diferentes)

75 términos no preferentes (USE)

17 relaciones de especificidad (NA)

1490 relaciones jerárquicas (TA/TE)

1074 relaciones asociativas (TR)

El Tesoro de Educación Superior de la Universidad Complutense puede consultarse en:

http://www.ucm.es/info/DAP/pr35/pr35_entrada.htm

Edición impresa:

Tesoro de Educación Superior / Gimeno Perelló, Javier, coord.; Lucas Padín, Eladio; Sánchez Melero, Ana [et.al.]. - Madrid: Universidad Complutense. Departamento de Análisis y Planificación, 1999.

BIBLIOGRAFÍA

- ?? Foucault, M.: **La arqueología del saber**. México: Siglo XXI, 1990
- ?? García Gutiérrez, A.L.: **Estructura lingüística de la documentación : teoría y método**. Murcia: Universidad, 1990
- ?? Herrero Solana, V.F.: **Hiperdocumentos referenciales**. Buenos Aires: Nuevo Paradigma, 1998
- ?? Izquierdo Arroyo, J.M.: **La organización mental del conocimiento**. 2 t. en 1 v. Madrid: Tecnidoc, 1995
- ?? Kedrov, B.M.: **Clasificación de las ciencias**. Moscú: Progreso, 1976. 2 v.
- ?? Nelson, T.H.: *Interactive systems and the design of virtuality*. **Creative computer**, v.6, nº 11, nov. 1980
- ?? Perec, G.: **Pensar, clasificar**. Barcelona : Gedisa, 1986
- ?? **Principios directrices para la elaboración y desarrollo de tesauros monolingües**. ISO 2788-1986. Ginebra, 1986
- ?? San Segundo, R.: **Sistemas de organización del conocimiento : La organización del conocimiento en las bibliotecas españolas**. Madrid: Universidad Carlos III; BOE, 1996.

SOBRE EL AUTOR

Javier Gimeno Perelló es licenciado en Filología Hispánica, Master en Documentación científica y Doctorado en CC. de la Información por la Universidad Complutense de Madrid, y Especialista universitario en Ciencia, Tecnología y Sociedad por la Universidad Nacional de Educación a Distancia.

Miembro del cuerpo facultativo de bibliotecarios de España, ha dirigido bibliotecas universitarias, públicas y de investigación, e impartido cursos y seminarios de documentación científica y lingüística documental en

universidades y otros centros académicos. En la actualidad dirige el centro de documentación especializado en enseñanza superior de la Universidad Complutense de Madrid.

Ha trabajado en proyectos de cooperación al desarrollo en materia de biblioteconomía, información y documentación en países de América Latina y del Caribe, participado en congresos y conferencias y publicado artículos en revistas científicas y profesionales y libros sobre estas disciplinas, entre los que destacan: *Tesaurus de Lingüística*; *El Centro de documentación en comunicación social de la Universidad Centroamericana de Managua*; *Operaciones de la cadena documental* (en col.); *Tratado básico de biblioteconomía* (en col.); *Tesaurus de Educación Superior* (coord.); *La Universidad en el comienzo de siglo: una respuesta entre el pragmatismo y la utopía* (en col.