

**Universidad Complutense de Madrid**  
**FACULTAD DE CIENCIAS DE LA INFORMACIÓN**  
**DEPARTAMENTO DE BIBLIOTECONOMÍA Y DOCUMENTACIÓN**



**Evolución histórica de los tesauros españoles  
y análisis de su rendimiento en el proceso  
de recuperación de información**

**Trabajo de investigación que presenta Blanca  
Gil Urdiciain para la obtención del grado  
de doctora en Ciencias de la Información  
bajo la dirección del Prof. Dr.  
José López Yepes**

Se recuerda al lector no hacer más  
uso de esta obra que el que  
permiten las disposiciones Vigentes  
sobre los Derechos de Propiedad  
Intelectual del autor. La Biblioteca  
queda exenta de toda responsabilidad.

Dado de Baja  
en la  
Biblioteca

Madrid, 1997

**UNIVERSIDAD COMPLUTENSE  
DE MADRID**  
**FACULTAD DE CIENCIAS  
DE LA INFORMACION**  
**REGISTROS DE LIBROS**  
**BIBLIOTECA GENERAL**  
Nº Registro ..... J.D. 476 .....

**Sic vos non vobis**

## **Agradecimientos**

Deseo expresar mi reconocimiento a todas las personas que de un modo u otro han contribuido a la realización de esta tesis doctoral. Así, quiero agradecer a Angel Villagrà su apoyo, consejo y comprensión en los decisivos momentos del comienzo de este trabajo y en sus fases posteriores; a Amelia de Irazazábal y a José María Sánchez Nistal su paciencia para reconstruir datos relativos a los orígenes del tesoro. A mis compañeros Fernando Ramos y Juan Gracia, por sus ánimos en los momentos más bajos. A Sara Angulo, Paquita Pulido y David Aceituno, que me ayudaron a comprobar la relevancia de los documentos recuperados y a mi familia, mi mejor y más paciente aliada a la hora de llevar a cabo la tediosa y mecánica tarea de confrontar los datos resultantes de cada consulta.

Mi agradecimiento igualmente a las personas que atendieron las complejas demandas en las bases de datos sometidas a análisis. Asimismo, deseo agradecer a Clara Giner, Laura Barrios y, de forma especial, a José Rojo, del Centro de Cálculo del CSIC, su ayuda en la preparación de la muestra aleatoria que resultó imprescindible para concluir este trabajo.

Quiero, por último, resaltar el apoyo prestado por el profesor López Yepes, mi director de tesis, sin cuyo estímulo y confianza, este trabajo no se hubiera llevado a cabo.

# Índices

## Índice general

<b>Capítulo 1. Objetivos, método, fuentes y bibliografía .....</b>	<b>10</b>
1.1. Objetivos y alcance de la investigación .....	11
1.1.1. Estructura y organización de la investigación .....	13
1.2. Método .....	15
1.3. Fuentes y bibliografía .....	17
1.4. Hipótesis generales .....	23
<b>Capítulo 2. El tesoro en el marco de los lenguajes documentales .....</b>	<b>25</b>
2.1. Introducción .....	26
2.2. Concepto de lenguaje documental .....	27
2.3. Características del lenguaje documental .....	29
2.3.1. Funciones del lenguaje documental .....	30
2.4. Tipología de los lenguajes documentales .....	32
2.5. Concepto de tesoro documental .....	35
2.5.1. Funciones del tesoro .....	37
2.5.2. Composición del tesoro .....	40
2.5.3. Morfología de los descriptores .....	49
2.5.4. Relaciones entre los términos del tesoro .....	52
2.5.4.1. Relaciones jerárquicas o paradigmáticas .....	53
2.5.4.2. Relaciones asociativas o sintagmáticas .....	56
2.5.4.3. Relaciones de equivalencia .....	59
2.5.5. Formas de presentación de los descriptores .....	59
2.5.5.1. Presentación alfabética .....	61
2.5.5.2. Presentación sistemática .....	63
2.5.5.3. Presentación gráfica .....	65
2.5.5.3.1. Diagramas arborescentes .....	68
2.5.5.3.2. Diagramas de flechas .....	68
2.5.5.3.3. Terminogramas .....	71
2.5.5.4. Índices auxiliares .....	72
2.6. Mantenimiento y actualización de tesauros .....	75
<b>Capítulo 3. Orígenes y evolución de los tesauros en España .....</b>	<b>84</b>
3.1. Antecedentes históricos del tesoro .....	85
3.2. Tesauros documentales .....	99
3.2.1. Formas alternativas de organización de los tesauros .....	104
3.3. Desarrollo de tesauros documentales en España .....	106
3.3.1. Introducción .....	106
3.3.2. Los tesauros en el ámbito formativo .....	107
3.3.3. Los primeros tesauros españoles .....	119
3.3.3.1. Otras referencias y proyectos de tesauros .....	124

3.4. Tesoros elaborados en España .....	147
3.4.1. Tesoros en fase de elaboración .....	163
3.4.2. Tesoros internacionales con versión o traducción en español, catalán o valenciano .....	166
<b>Capítulo 4. La recuperación de información .....</b>	<b>176</b>
4.1. Introducción .....	177
4.2. Sistemas de Información .....	178
4.3. Bases de datos .....	184
4.3.1. Registros de una base de datos .....	186
4.4. Búsqueda documental .....	199
4.4.1. Análisis de los conceptos de la consulta .....	199
4.4.2. Elección del tipo de lenguaje a utilizar .....	200
4.4.3. Definición de la estrategia de búsqueda .....	201
4.4.4. Establecimiento de enlaces sintácticos entre los términos de búsqueda .....	203
4.4.4.1. Álgebra de Boole .....	203
4.4.4.2. Truncamiento .....	206
4.4.4.3. Enlaces de proximidad .....	208
4.4.4.4. Método de ponderación o recuperación por pesos .....	209
4.4.4.5. Método estadístico .....	212
4.4.4.6. Recuperación mediante tesoro .....	213
4.4.5. Visualización y evaluación de los resultados .....	215
4.4.6. Exportación de los registros .....	216
<b>Capítulo 5. Análisis semántico y estructural de diferentes tesoros y características de las bases de datos en las que se utilizan .....</b>	<b>218</b>
5.1. Introducción .....	219
5.2. Criterios de evaluación .....	220
5.3. Características de los tesoros analizados .....	224
5.3.1. Tesoro ICYT de Biología animal .....	224
5.3.2. Tesoro Electrotécnico .....	227
5.3.3. Tesoro de Medio ambiente .....	233
5.3.4. Tesoro "Mujer" .....	237
5.3.5. Tesoro ISOC de Psicología .....	241
5.3.6. Tesoro de Asuntos sociales .....	246
5.4. Características de las bases de datos consultadas .....	249
5.4.1. Base de datos ICYT de Biología animal .....	249
5.4.2. Base de datos PIE (Programa de Investigación y Desarrollo Electrotécnico) .....	251
5.4.3. Base de datos Biblioma del Centro de Documentación del Ministerio de Medio Ambiente .....	257
5.4.4. Base de datos del Centro de Documentación de la Mujer .....	259
5.4.5. Base de datos PSEDISOC del CSIC .....	263
5.4.6. Base de datos de la Biblioteca del Instituto de Migraciones y Servicios sociales .....	265

<b>Capítulo 6. Evaluación del rendimiento de tesauros españoles en sistemas de recuperación de información .....</b>	<b>270</b>
6.1. Introducción .....	271
6.2. Antecedentes .....	273
6.3. Metodología .....	280
6.4. Desarrollo de las búsquedas .....	285
6.4.1. Aspectos y consideraciones para la preparación de las estrategias de búsqueda .....	287
6.5. Criterios de evaluación .....	290
6.6. Desarrollo de las estrategias .....	295
6.6.1. Base de datos ICYT de Biología animal .....	295
6.6.2. Base de datos PIE (Programa de Investigación y Desarrollo Electrotécnico) .....	309
6.6.3. Base de datos Biblioma, del Ministerio de Medio Ambiente .....	323
6.6.4. Base de datos del Centro de Documentación de la Mujer .....	338
6.6.5. Base de datos PSEDISOC, del CSIC .....	353
6.6.6. Base de datos de la Biblioteca del Instituto de Migraciones y Servicios sociales .....	367
<b>Capítulo 7. Conclusiones .....</b>	<b>387</b>
<b>Bibliografía .....</b>	<b>394</b>

## **Índice de figuras**

Figura 1: Tipología de los lenguajes documentales en función del control ejercido sobre su vocabulario.

Figura 2: Tipología de los lenguajes documentales según el criterio de coordinación.

Figura 3: Tipología de los lenguajes documentales en función de su estructura.

Figura 4: El papel del tesoro en un sistema de documentación (Tesoro Spines).

Figura 5: Composición del tesoro

Figura 6: Tipología del descriptor.

Figura 7: Extracto del índice alfabético estructurado del Tesoro ISOC de Psicología.

Figura 8: Extracto del índice sistemático del Tesoro ISOC de Psicología.

Figura 9: Diagrama arborescente. Tesoro URBAMET.

Figura 10: Diagrama flechado de descriptores. Thesaurus International Road Research Documentation (IRRD).

Figura 11: Terminograma de descriptores. Tesoro de Servicios sociales del Inersro.

Figura 12: Extracto del índice permutado del Tesoro ISOC de Psicología.

Figura 13: Modelo para el registro de descriptores

Figuras 14 y 15: Thesaurus Linguae Latinae

Figura 16: English Synonymes Explained in Alphabetical Order with copious ilustrations and examples.

Figuras 17 y 18: Roget´s Thesaurus.

Figura 19: LEDA Thesaurus.

Figuras 20, 21 y 22: Planes de estudios (publicación BOE) de las Escuelas de Biblioteconomía y Documentación de las Universidades de Granada, Barcelona y Salamanca.

Figuras 23 y 24: Programa del Seminario sobre La Documentación al servicio de la industria química, farmacéutica e investigación científica.

Figuras 25 y 26: Programa de Seminario sobre Documentación científica.

Figura 27 y 28: Programa de Cursillo de Documentación química.

Figuras 29, 30 y 31: Thesaurus Textil Internacional.

Figura 32: Tesoro de Materias grasas y derivados.

Figuras 33 y 34: Extracto de una publicación del Instituto Nacional de Seguridad e Higiene en el Trabajo.

Figuras 35 y 36: Nomenclatura codificada de diagnósticos.

Figuras 37 y 38: Tesoro ISO. IRANOR.

Figuras 39, 40 y 41: Primeras pantallas del menú de las bases de datos del CSIC.

Figuras 42, 43 y 44: Primeras pantallas del menú de la base de datos PIE.

Figuras 45, 46 y 47: Primeras pantallas del menú de la base de datos del Centro de Documentación de la Mujer.

Figuras 48 y 49: Primeras pantallas del menú de la base de datos de la Biblioteca del Instituto de Migraciones y Servicios sociales.

# Capítulo 1

## Objetivos, método, fuentes y bibliografía

## **1.1. Objetivos y alcance de la investigación**

El propósito general de esta tesis doctoral es ofrecer una panorámica del desarrollo y evolución histórica de los tesauros en España y evaluar su eficacia en el proceso de recuperación de información.

En relación con el primer objetivo, se perfilan los antecedentes, orígenes e historia del tesoro. Por lo que se refiere a la evaluación del rendimiento, la tesis presenta un trabajo experimental realizado en bases de datos reales, cuyos resultados servirán de fundamento para poner de manifiesto las ventajas e inconvenientes del empleo de tesauros en centros de documentación y/o bibliotecas contemporáneas, así como sus posibilidades como útil al servicio de la recuperación de información en un futuro próximo. Así pues, el objeto de análisis, el tesoro, se sitúa en el contexto del proceso documental en torno a las operaciones de representación y búsqueda documentales.

El tipo de investigación empírica que proponemos ha sido desarrollada en otros países con diversas aproximaciones al tema. Sin embargo, en España no se han publicado valoraciones de las prestaciones de los distintos lenguajes en el momento de la búsqueda informativa, y menos aún, en lo que se refiere al lenguaje documental aquí tratado.

Elegimos como tema de estudio el tesoro y sus características de empleo porque, desde nuestro punto de vista de profesionales de la Documentación, teníamos una vaga idea de la situación existente en relación con el tema y nos pareció importante conocer la situación actual en lo que a utilización y prestaciones de este lenguaje documental se

refiere. De los resultados de un trabajo anterior<sup>1</sup> deducíamos que, en la categoría de tesaurus, se englobaban realidades que no ofrecían las peculiaridades de este lenguaje de recuperación de información. Por otra parte, nos animó el hecho de que, aunque existe abundante literatura acerca de las ventajas y desventajas del uso del lenguaje libre frente al controlado y viceversa, esta información no se había contrastado en nuestro país. El interés mostrado por especialistas del CSIC en materia de análisis y lenguajes documentales fue otro aliciente que nos decidió firmemente a seguir esta línea de investigación.

A pesar de que, en respuesta al desarrollo exponencial de la producción bibliográfica de la últimas décadas, han surgido una proliferación de medios informáticos para canalizarla, el tesaurus parece ser la solución propuesta por los documentalistas a los problemas de recuperación del conocimiento.

Las dificultades de localización de datos se han visto multiplicadas con el desarrollo de redes como, por ejemplo, Internet, potencialmente fuente inagotable de datos, por el hecho de que la información que fluye por la red no está sometida a ningún tipo de control terminológico, viéndose menguadas sus posibilidades de explotación. El gasto de tiempo, entre otros inconvenientes, es consecuencia de esa falta de control del *vocabulario empleado, tanto para alimentar el sistema como para buscar información de interés*. Si los procesos de recuperación de información son complejos incluso para especialistas en el manejo de técnicas de tratamiento documental, la dificultad se acentúa cuando se trata de personas que desconocen dichas técnicas y que se enfrentan además a sistemas que no ofrecen ningún género de normalización en la información que gestionan. Los documentalistas diseñan los sistemas dando por supuesto que los usuarios van

---

<sup>1</sup>GIL URDICIAIN, B. Tesaurus, en: **Manual de lenguajes documentales**. Madrid: Noesis, 1996.

a saber utilizarlos y esto no es siempre así.

Las técnicas documentales se han empleado tradicionalmente como recurso adecuado para poder seleccionar la información y poner a disposición del usuario los datos por él solicitados. Todas estas técnicas tienen el común denominador de la normalización, que hace viable la recuperación a nivel individual y en red. Nosotros apostamos por su empleo sistemático, especialmente por el uso de un lenguaje controlado que, combinado con las crecientes posibilidades de las aplicaciones informáticas, consideramos capaz de manejar la multiplicidad de datos que se van generando, incluso en el entorno de los más ambiciosos programas de la industria de la información.

En síntesis, el objetivo general de la investigación comporta el estudio de las siguientes cuestiones:

- Presentar al tesoro como útil al servicio de la recuperación de información.
- Averiguar los orígenes de los tesauros en España.
- Realizar un análisis semántico y estructural de diferentes tesauros.
- Evaluar la eficacia de los tesauros en el proceso de recuperación documental.

### **1.1.1. Estructura y organización de la investigación**

Se trata de una investigación organizada en dos planos, un primer plano de investigación histórica, y el segundo, de carácter experimental. El trabajo se divide en siete capítulos, seis de ellos monográficos, destinados a profundizar desde diferentes enfoques en los aspectos más significativos de los temas de estudio.

En este primer capítulo **Objetivos, método, fuentes y bibliografía**, se presentan los elementos generales componentes de la tesis. Se plantea el tema central del trabajo con sus objetivos y alcance, el método general y la organización de sus partes componentes, las hipótesis generales, las fuentes y la bibliografía consultadas.

El capítulo segundo, **El tesoro en el marco de los lenguajes documentales** está dedicado íntegramente a presentar al tesoro como lenguaje documental, es decir, como *elemento normalizador en los procesos de indización y recuperación documentales*, operación, esta última, en la que nos basamos para realizar el trabajo experimental que presentamos en el penúltimo apartado de esta tesis doctoral.

El tercer capítulo, **Orígenes y evolución de los tesauros en España** ofrece un análisis global de la evolución del término *thesaurus* y de los orígenes de su empleo *como útil al servicio de la recuperación de información*. La segunda parte del capítulo describe el panorama general de la producción de tesauros en España e incluye un epígrafe con el censo de los elaborados y en circulación en la actualidad en nuestro país.

En el capítulo cuarto, **La recuperación de información**, se describe esta fase clave en el tratamiento documental y la que justifica todas las demás. Se presentan en este apartado los componentes de un sistema de recuperación de información, así como métodos alternativos de búsqueda, muchos de los cuales fueron puestos en práctica para el desarrollo del test, objeto de esta investigación.

El quinto capítulo, **Análisis semántico y estructural de diferentes tesauros y características de las bases de datos en las que se utilizan**, se destina al análisis de los componentes de cada una de las bases de datos consultadas para llevar a cabo el trabajo

experimental, y la evaluación de las unidades léxicas y las características estructurales de seis tesauros utilizados en dichas bases de datos.

El sexto capítulo, **Evaluación del rendimiento de tesauros españoles en sistemas de recuperación de información**, constituye uno de los núcleos de esta tesis doctoral. Se realiza en esta parte un exhaustivo análisis del rendimiento de los tesauros empleados por seis bases de datos en el momento de recuperar la información. Para valorar los resultados se tienen en cuenta la precisión y exhaustividad, parámetros validados por la comunidad científica.

El trabajo se cierra, como es preceptivo, con un capítulo de **Conclusiones**, que constituye una síntesis de los elementos concluyentes derivados de los resultados obtenidos en la investigación.

## **1.2. Método**

Para realizar esta investigación se han empleado dos métodos completamente distintos, en virtud de los dos aspectos estudiados. La perspectiva histórica, que conforma el *primer frente de investigación*, se encuadra, *panorámicamente*, en el marco metodológico del análisis histórico. Se procedió fundamentalmente a realizar una aproximación descriptiva de las fuentes documentales halladas.

Por lo que se refiere al trabajo experimental, se realizó en seis bases de datos distribuidas, a partes iguales, en las áreas de ciencias sociales y tecnología, siguiendo el objetivo final de que los resultados de la investigación fueran extrapolables a otras bases de datos, independientemente de su composición. Se procedió mediante la combinación

de los métodos analítico y el de muestreo y validación de datos. En la fase de muestreo se tendió a maximizar la amplitud y calidad de la información recogida a fin de medir los valores cuantitativos *precisión* y *exhaustividad*, que habían de tenerse en cuenta para establecer el análisis comparativo.

La fase de recopilación de datos se desarrolló durante un período de diez meses. Los datos se imprimieron en algunos centros de documentación y bases de datos, en otros, se recogieron mediante grabación en disquete. En ambos casos, en una fase posterior, se contrastó la información resultante con las estrategias de búsqueda propuestas.

En el proceso de validación, adoptamos la metodología propuesta por Cleverdon<sup>2</sup> en un estudio pionero de las características del que aquí llevamos a cabo, y que ha venido siendo utilizado por otros investigadores del área con excelentes resultados. Dicho proceso fue laborioso y se centró en el comportamiento de los tesauros y el lenguaje libre en base a cada una de las alternativas de búsqueda que se hicieron con cada consulta.

Hallar los índices de precisión no fue tarea complicada dado que su cálculo requiere el conocimiento de datos fácilmente contrastables. Valorar la exhaustividad, en cambio, supuso un gran esfuerzo ya que, hallar este parámetro, supone el conocimiento previo de todos los documentos relevantes de un fondo documental que respondan a determinado tema de búsqueda. A excepción de una base de datos, que cuenta con un volumen restringido de documentos, este fue un valor materialmente imposible de conocer, por lo que se hizo necesario proceder mediante métodos alternativos, entre

---

<sup>2</sup>CLEVERDON, C.W. *ASLIB Cranfield research project: report on the testing and analysis of an investigation into the comparative efficiency of indexing systems*. Cranfield: College of Aeronautics, 1962.

otros, técnicas de muestreo que se pusieron en práctica previa generación de números aleatorios, y manejando el factor varianza.

Queremos advertir que, en todo momento, partimos del conocimiento de que todo trabajo de investigación se halla mediatizado por elementos propios del contexto que, al ser sumamente específicos, inciden de forma concreta en los resultados obtenidos. En nuestro caso, el contexto desempeñó un papel activo sobre la información recuperada. Hubiera sido ideal poder aislar todas las posibles variables: características de la indización, normalización de los medios de alimentación y recuperación de las bases de datos consultadas, etc., pero el control de dichas variables contextuales hubiera supuesto un cambio total en los planteamientos de esta tesis, algo que queríamos evitar. No deseábamos realizar un estudio de laboratorio, por lo que desde un principio nos pareció oportuno obtener resultados de búsquedas reales, no sometidas a previo control.

La utilización de una misma metodología en todos los casos analizados, requirió el trabajo previo de resolver problemas derivados de la escasez de centros de documentación o bibliotecas provistas de tesoro propio, así como de la delimitación de potenciales temas de consulta que permitieran preparar estrategias de búsqueda con un cierto común denominador en todas las bases.

### **1.3. Fuentes y bibliografía**

La documentación consultada tuvo dos características bien diferenciadas, de acuerdo con los dos aspectos fundamentales que componen esta investigación: por una parte, el estudio de los orígenes de los tesauros en España, por otra parte, la evaluación de su rendimiento en sistemas de recuperación de información.

En relación con los antecedentes históricos del tesoro en España, podemos señalar la inexistencia de datos anteriores a 1969, lo cual hizo necesaria la incidencia en fuentes informales, en el más estricto sentido, siendo las bibliográficas válidas sólo a partir de la fecha señalada. Quiero señalar aquí la dificultad de acceso a las fuentes documentales por su escasez, la ambigüedad y la calidad de inéditas de las mismas. La localización de literatura gris, determinados informes en concreto, fue definitiva para poder materializar información significativa en relación con la producción de lenguajes documentales en nuestro país, habiendo tenido que reconstruir datos acudiendo al testimonio directo de los creadores de algunos tesauros.

Para la recopilación de antecedentes históricos se procedió al vaciado de **bibliografías** internacionales, publicadas entre los años 1975 y 1988, en las cuales no apareció referencia alguna a tesauros preparados en España.

La consulta de los **planes de estudios** de las primeras diplomaturas en Biblioteconomía y Documentación fue de utilidad, al igual que el análisis de los *programas de los primeros cursos impartidos sobre Documentación en diferentes instituciones*, tales como La Asociación Nacional de Químicos. También fue de interés la consulta del Proyecto de Plan de estudios sobre Documentación científica para *postgraduados universitarios, preparado por la doctora Emilia Currás en 1977.*

Por lo que se refiere a los restantes temas de estudio en este trabajo, se procedió al análisis y revisión bibliográficas de **monografías nacionales e internacionales**, cuyo contenido trataba parcialmente dichos temas. Este tipo de obras fue útil, particularmente, para los capítulos referidos a los fundamentos teóricos del tesoro, la recuperación de información, la evaluación semántica de los tesauros y la evaluación de la eficacia de los lenguajes documentales en la recuperación de información. En función de los datos que

nos aportaron sobre las mencionadas materias, los títulos más representativos fueron respectivamente: Van Slype: *Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales* (1991); la monografía de Miguel Castaño y Piattini Velthuis: *Concepción y diseño de bases de datos: Del modelo E/R al modelo relacional* (1993), que nos sirvió de fundamento para el capítulo referido a la recuperación documental al ofrecer amplia información acerca de los servicios de información. Lancaster, en su obra: *El control del vocabulario en la recuperación de información* (1995) propone, en el capítulo 16, una metodología para la evaluación semántica y estructural de los tesauros, en la que nos basamos para realizar el análisis de dichos lenguajes documentales en esta investigación. Es importante citar la obra más completa, a nuestro juicio, sobre búsqueda de información en bases de datos: *La búsqueda documental en el contexto telemático: Modalidades de automatización y utilización de bases de datos*, preparada por Bares y traducida al español en 1988. El libro escrito en colaboración por Salton y McGill: *Introduction to modern information retrieval* (1983) contribuye en gran medida a comprender la importancia del tesoro en los procesos de recuperación de información. La consulta de esta fuente también fue de utilidad en el momento de plantear el procedimiento más adecuado de evaluación del rendimiento de los tesauros.

Debemos hacer mención especial a la obra de Cleverdon: *ASLIB Cranfield research project: report on the testing and analysis of an investigation into the comparative efficiency of indexing systems*, publicada en 1962. El autor propone una metodología de evaluación de lenguajes documentales, que sirvió de modelo a la que proponemos en el sexto capítulo de esta tesis doctoral. Dicha metodología nos fue de utilidad para hallar los índices de *precisión* y *exhaustividad* de los documentos recuperados y comparar, de este modo, la eficacia de los tesauros. Tenemos que decir, sin embargo, que la citada obra describe un test realizado en condiciones de laboratorio, con control previo de todas

las variables que intervienen en los procesos de tratamiento y recuperación documentales. En nuestro caso, la experimentación se llevó a cabo en bases de datos reales, sin previo control de dichas variables, de modo que las orientaciones metodológicas del *ASLIB Cranfield research project* sólo nos sirvieron parcialmente.

En cuanto a las **fuentes hemerográficas**, se procedió al análisis y vaciado de revistas españolas y extranjeras sobre Biblioteconomía y Documentación, que no siempre fueron de utilidad. Si nos remitimos al primer aspecto tratado en esta investigación, es decir, el aspecto histórico, podemos confirmar la inexistencia de datos al respecto en publicaciones periódicas, independientemente de su procedencia. Citaremos a continuación algunos artículos que tratan de los orígenes de los lenguajes documentales y de los tesauros, pero que, en ningún caso, hacen mención a su desarrollo en España: el trabajo de Sparck Jones: *Some thesauric history*, *Aslib Proceedings*, 7 (1972); la recopilación preparada por Roberts: *The pre-history of the information retrieval thesaurus*, *Journal of Documentation*, 4 (1984); *Los tesauros de términos científicos y técnicos en general, y especialmente los relativos a la química (revisión bibliográfica)*, trabajo publicado por Sancho Lozano en la *Revista Española de Documentación Científica* (1985); por último, el artículo preparado por López Huertas Pérez: *Lenguajes documentales: aproximación a la evolución histórica de un concepto*, publicado en el *Boletín de ANABAD*, 1 (1991).

En lo que al segundo frente de investigación se refiere, se tuvieron en cuenta todos aquellos artículos que se centraban en el análisis y evaluación de lenguajes documentales. Fueron numerosas las publicaciones que hacían referencia a tests comparativos de la eficacia del lenguaje libre y el controlado, pero no describían experimentaciones desarrolladas en bases de datos reales, o bien se planteaban a niveles de complejidad que no nos era posible abarcar en un trabajo de carácter individual como el que aquí proponemos. De las fuentes consultadas podríamos destacar las siguientes: *Information*

*retrieval experimentation*, artículo de gran interés preparado por Tague-Sutcliffe e incluido en la *Encyclopedia of Library and Information Science*. En él se señalan los tests más importantes realizados para evaluar lenguajes y sistemas de recuperación de información, así como las medidas empleadas para efectuar dicha valoración. El Journal of Documentation publicó, en marzo de 1969, el trabajo firmado por Robertson: *The parametric description of retrieval tests*, cuya importancia reside en la minuciosa explicación que ofrece de los parámetros y técnicas empleadas para describir los resultados de evaluaciones de lenguajes documentales y servicios de recuperación de información. En marzo de 1973, el Journal de Documentation publicó el artículo que bajo el título: *The Aberystwyth index languages test*, describía el test realizado por M. Keen, en el College of Librarianship Wales. En el citado test se compara la efectividad de cinco lenguajes de indización. La metodología propuesta nos fue de utilidad, con limitaciones, dado que, al igual que en la experimentación realizada por Cleverdon, se planteaba un estudio de laboratorio.

También se realizaron consultas en **Bases de datos nacionales e internacionales**. El tipo de consulta realizada en las mismas fue de características bien distintas, las cuales detallamos a continuación:

*Base de datos Thesauri*, del distribuidor ECHO, perteneciente a las Comunidades Europeas. En esta base se solicitaron datos sobre los tesauros españoles existentes. Lamentablemente, la respuesta fue nula, puesto que no tenían registrado ni un sólo tesaurus elaborado en nuestro país.

*Base de datos LISA (1991-1997)*. Su consulta proporcionó datos bibliográficos de interés, básicamente, para investigar en los orígenes de los tesauros, en general, pero no aportó información de utilidad en relación con los construidos en España.

*Base de datos ICYT de Biología animal*

*Base de datos PIE*

*Base de datos Biblioma*

*Base de datos del Centro de Documentación de la Mujer*

*Base de datos PSEDISOC del CSIC*

*Base de datos de la Biblioteca del Instituto de Migraciones y Servicios sociales*

La consulta de estas seis bases no tuvo su origen en la búsqueda de información relevante sobre los distintos temas tratados en esta tesis doctoral, sino en la realización del test de carácter experimental que proponemos en el sexto capítulo. Las búsquedas se hicieron, pues, de forma exhaustiva y las materias de consulta nos sirvieron, en tanto que instrumento para la comparación de los lenguajes libre y controlado que realizamos en virtud de los resultados obtenidos.

En cuanto a las principales bibliotecas consultadas fueron: la *Biblioteca Nacional*, la *Biblioteca del Bibliotecario*, las *Bibliotecas del ISOC y del ICYT*, el *Centro de documentación de Fundesco*, la *Biblioteca de la Escuela Universitaria de Biblioteconomía y Documentación, de la Universidad Complutense* y, por supuesto, los centros de documentación y bibliotecas en las que se encontraban las bases de datos que acabamos de mencionar, en las que se desarrolló la experimentación.

Por lo que se refiere a la información bibliográfica que aportamos al final de este trabajo, hemos tratado de ser exhaustivos, limitándonos, no obstante, a aquellas obras que, de forma más o menos directa, han servido para la realización de esta tesis doctoral.

Para facilitar su consulta, dicho repertorio bibliográfico se ha estructurado sistemáticamente de acuerdo con los aspectos que conforman la investigación y, con ese

mismo objetivo, se ha organizado siguiendo las normas internacionales de descripción bibliográfica ISBD para el tipo de material referenciado: monografías y publicaciones seriadas.

#### **1.4. Hipótesis generales**

Como fase previa a la realización del trabajo hemos fijado -tratando de documentarlas a posteriori- las siguientes hipótesis:

- El índice de utilización de tesauros en bases de datos españolas es bajo debido, bien a la falta de adecuación de dichos lenguajes documentales a necesidades específicas, o simplemente a su inexistencia en áreas temáticas especializadas.

- El experimento tiene considerables limitaciones, especialmente debidas a la imposibilidad de controlar todas las variables del proceso documental, pero presuponemos que el control del vocabulario es factor determinante en la última fase de dicho proceso, para conseguir altos niveles de precisión informativa, independientemente del método de búsqueda que se practique.

- El uso de un lenguaje controlado facilita la búsqueda documental, tanto en bases de datos técnicas como en aquellas que desarrollan su actividad en el campo de las ciencias sociales y humanidades.

- La recuperación mediante lenguaje libre genera un elevado número de documentos irrelevantes.

- La consulta de un tesoro, en el momento de preparar las estrategias de búsqueda, supone menor esfuerzo para el usuario porque le evita el esfuerzo de pensar en todas las posibles acepciones y relaciones semánticas que pueda tener un término tomado del lenguaje natural. En este segundo caso, es el propio consultante quien ha de realizar un control previo del vocabulario para proceder a la recuperación.

## Capítulo 2

### El tesoro en el marco de los lenguajes documentales

## **2.1. Introducción**

En este capítulo presentamos las características, funciones y tipología del lenguaje documental, en general, y del tesoro en particular, como elemento normalizador en los procesos de tratamiento y recuperación documentales.

Los documentos que componen una unidad de información o biblioteca son analizados por los profesionales del documento para hacer posible su consulta por el usuario. Este proceso, denominado Análisis documental, tiene por objeto dotar a cada libro de una serie de puntos de acceso, para permitir su posterior recuperación por cualquiera de ellos. Estos puntos de acceso pueden consistir en el nombre del autor de la obra, de su título o del tema que trate. Del análisis de los aspectos externos de un documento se pueden tomar los datos relativos a su título y autor, lo que constituye la descripción bibliográfica. El tema, los conceptos que implícita o explícitamente están recogidos en una obra se extraen mediante el Análisis de Contenido, cuyas operaciones: indización y resumen, dan al lector un conocimiento más o menos profundo de la información que contiene. La operación de indizar consiste, en efecto en el análisis e identificación de los conceptos del documento y la selección de aquellas nociones que representen con mayor fidelidad la información que contiene. Para normalizar la denominación de dichos conceptos se procede a su traducción a un lenguaje documental, lo cual facilita la recuperación, independientemente del analista que trate el documento.

## 2.2. Concepto de lenguaje documental

La principal característica de los lenguajes documentales es su capacidad de control terminológico, que permite la expresión de los conceptos derivados de la indización sin ambigüedades, garantizando el empleo de los mismos términos para representar los mismos conceptos. La definición de Van Slype refleja el componente pragmático del lenguaje documental: *todo sistema de signos que permita representar el contenido de los documentos con el fin de recuperar los documentos pertinentes en respuesta a consultas que tratan sobre ese contenido*<sup>1</sup>.

García Gutiérrez concibe el lenguaje documental desde la perspectiva de la teoría de la comunicación, destacando su aspecto normativo: *conjunto normalizado y normativo de términos relacionados por principios comunes, declarados portavoces preferenciales de los mensajes encerrados en un colectivo documental con el fin de provocar una recuperación pertinente de información por aproximación temática*<sup>2</sup>.

Nosotros entendemos por lenguaje documental *todo sistema artificial de signos normalizados, que facilitan la representación formalizada del contenido de los documentos para permitir la recuperación, manual o automática, de información solicitada por los usuarios.*

Las primeras manifestaciones del lenguaje documental datan de finales del siglo XIX, cuando aparecen las clasificaciones bibliográficas. Estas clasificaciones, inspiradas

---

<sup>1</sup>SLYPE, G.v. **Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales.** Madrid: Fundación Germán Sánchez Ruipérez, 1991, p. 21.

<sup>2</sup>GARCÍA GUTIÉRREZ, A. **Lingüística documental: aplicación a la documentación de la comunicación social.** Barcelona: Mitre, 1984, p. 157.

en la lógica y en los sistemas filosóficos del conocimiento, se fundan en el principio de *precoordinación* y son de carácter enciclopédico.

El concepto moderno se consolidó en el siglo XX cuando Cutter introdujo el lenguaje de encabezamientos de materia, basado en los principios de especificidad y de entrada directa, que señala el comienzo del desarrollo de lenguajes documentales especializados, entre los que se encuentra el tesoro, objeto de estudio en esta tesis doctoral.

Los lenguajes especializados, pues, nacieron como respuesta a una creciente especialización de los conocimientos (*derecho, economía, sociología, etc...*) que no podían ser representados mediante las materias contenidas en los lenguajes multidisciplinares existentes hasta entonces. Como consecuencia, se crearon numerosas clasificaciones especializadas, por ejemplo *Excerpta Medica*, así como múltiples tesauros sectoriales, con los que se podían organizar las colecciones a escala institucional, pero, al propio tiempo, provocaron *una confusión babélica a escala mundial*.<sup>3</sup> Parece, sin embargo, que existe actualmente una nueva tendencia hacia el enciclopedismo temático.

Pasamos a analizar en los siguientes epígrafes las partes componentes del lenguaje documental, así como sus principios y funciones en el proceso de tratamiento de la documentación.

---

<sup>3</sup>RODRÍGUEZ DELGADO, R. La integración de los lenguajes documentarios, fin de Babel. *Revista Española de Documentación Científica*, v. 4, n. 3 (1980).

### 2.3. Características del lenguaje documental

El lenguaje documental es un lenguaje no natural, aunque utiliza los signos de éste. Esos signos adquieren valor semántico por medio de su normalización y de las reglas morfosintácticas que lo articulan. Son precisamente esas reglas las que le dan categoría de lenguaje.

Los signos del lenguaje natural son las palabras, que representan nuestro conocimiento de la realidad; el lenguaje documental se sirve de ellas y, en ocasiones, las reemplaza por símbolos cargados de significado preciso de forma que lleva a cabo la representación del documento en virtud de una correspondencia analógica. Esta representación convencional tiene lugar cuando simboliza los elementos constitutivos del lenguaje de descripción (metalenguaje) por medio de instrumentos visuales de representación -la notación simbólica-; este es el caso de los sistemas de clasificación, que por el procedimiento de notación representan ideas de manera unívoca a través de números ordinales.

En comparación con el crecimiento de las ideas, el desarrollo del lenguaje natural es lento y, como consecuencia, es necesario utilizar la misma palabra para expresar dos o más ideas. Ello da lugar a la homonimia: el mismo término se utiliza en diferentes sentidos en diferentes disciplinas, lo que produce distorsiones en la comunicación. La sinonimia es otro accidente lingüístico habitual en el lenguaje documental, que debe ser reconocido y, o bien eliminado o, preferentemente, controlado.

Derivados de esos accidentes surge la ambigüedad, que crea dificultades estrechamente relacionadas con la aplicación de un lenguaje artificial a la recuperación de documentos. Se trata del ruido y silencio documentales, que se oponen a la relevancia

de la indización. Cuando se prepara una frase documental con objeto de recuperar un conjunto de documentos que contienen la información expresada en la frase, aquellos que no se recuperan, aunque existan en el archivo, son los productores del silencio documental, mientras que los que se recuperan sin haber sido solicitados -porque responden sólo parcialmente al perfil de búsqueda- constituyen el ruido documental. Sólo el logro de la deseable entropía puede evitar estos inconvenientes y facilitar una recuperación eficaz de la información.

### **2.3.1. Funciones del lenguaje documental**

El lenguaje documental interviene en dos fases del proceso documental: en el momento de la descripción y en el de la recuperación de la información. El objetivo de dichas operaciones es el de *facilitar la recuperación de la documentación reduciendo el esfuerzo y gasto de tiempo del usuario*. El lenguaje documental tiene capacidad para representar los mensajes contenidos en los documentos, lo que le permite cumplir dos *objetivos fundamentales en el proceso, el de normalización y el de inducción*, estando encaminadas a este último todas las demás funciones que desempeña a lo largo del proceso documental.

El lenguaje documental reduce considerablemente el volumen de términos del lenguaje natural no tomando en consideración más que los sustantivos o los sintagmas nominales, *seleccionando además, de entre ellos, un sólo término entre dos o más nociones consideradas sinónimas*. En virtud de estas características, se le reconoce un carácter mediador en los procesos de resumen e indización al controlar la posible ambigüedad entre términos.

En la fase de descripción, tras la lectura e identificación de los conceptos contenidos en los documentos, estos se representan mediante un lenguaje documental, que proporciona un vocabulario unívoco que favorece la coherencia del análisis y permite traducir los conceptos en términos normalizados.

Su objetivo primordial se cumple, sin embargo, en la segunda fase del proceso, suministrando el vocabulario de los conceptos de la demanda. El lenguaje documental proporciona una base para efectuar búsquedas al nivel de generalidad o especificidad deseados. En este sentido, se le puede considerar un lenguaje intermediario o metalenguaje en la medida en que sirve de puente entre las informaciones contenidas en los documentos y las informaciones solicitadas por los usuarios. Van Slype<sup>4</sup> considera que los lenguajes de indización pueden intervenir en seis momentos diferentes del proceso de búsqueda:

- Selección de los sistemas documentales que se van a interrogar
- Enunciado de los conceptos de la pregunta, en lenguaje natural
- Traducción a un lenguaje de indización
- Formulación de la ecuación
- Extensión asistida por el ordenador
- *Apreciación final de la pertinencia*

Además de las mencionadas funciones, el lenguaje documental es de gran utilidad para la ordenación o archivo de documentos. Resuelve igualmente problemas planteados por el multilingüismo, como en el caso de los tesauros multilingües, que sirven de puente a diferentes lenguas, o en el caso de los lenguajes de clasificación, fundados en la

---

<sup>4</sup>SLYPE, G. van. Op. cit., p. 161.

simbolización numérica, que los hace independientes de toda lengua y constituyen una escritura comprensible en todos los idiomas, de ahí su fácil empleo internacional.

#### 2.4. Tipología de los lenguajes documentales

La variada tipología del lenguaje documental le convierte en elemento de apoyo a disciplinas como la Biblioteconomía, Documentación, Bibliografía y Archivística, para cuyas necesidades de descripción ofrece posibilidades concretas. En relación con el Análisis formal, el lenguaje documental completa el proceso técnico de catalogación dotando al soporte de la descripción de puntos de acceso temáticos.

Existen diversos criterios de tipificación de los lenguajes documentales. Los más generalizados son los de **control**, el grado de **coordinación** de los términos y el **estructural**.

Dependiendo del **control** ejercido sobre el vocabulario, los lenguajes pueden organizarse en dos categorías: *libres* y *controlados* (Figura 1).

<b>Control</b>	<b>Libres</b>	Listas de descriptores libres
	<b>Controlados</b>	Clasificaciones, tesauros

Figura 1

Existe abundante literatura acerca de las ventajas y desventajas que conlleva el uso del lenguaje libre en las fases de tratamiento de la documentación. En esta tesis doctoral vamos a tratar de analizar comparativamente sus prestaciones frente a las que ofrece el tesauro. Nosotros, como ya explicábamos en la introducción del trabajo, apostamos por la normalización del vocabulario y esperamos poder aportar datos que permitan fundamentar esta opinión.

Los lenguajes libres, fundados en el principio de postcoordinación, se componen de un vocabulario no predefinido que se va generando a partir de la realización de procesos de indización. A excepción de las listas de descriptores libres y las de palabras clave, los demás lenguajes entran dentro de la categoría de controlados, que presentan un vocabulario previamente elaborado y que admite un limitado número de modificaciones en el momento de su utilización.

La sistematización de los lenguajes documentales según el criterio de **coordinación** (figura 2) se realiza en función del momento en que se combinan los elementos que los componen. Si los términos se combinan en el momento de la descripción, el lenguaje será *precoordinado*, y si lo hace en el momento de la recuperación, se tratará de un lenguaje *postcoordinado*. Van Slype, con un planteamiento pragmático, denomina lenguajes de clasificación a los *precoordinados* y lenguajes de indización a los *postcoordinados*; ambos tipos representan el contenido del documento de forma sintética y analítica, respectivamente.

<b>Coordinación</b>	<b>Precoordinados</b>	Clasificaciones
		Listas de encabezamientos de materia
	<b>Postcoordinados</b>	Listas de descriptores libres
		Listas de palabras clave
		Tesauros

Figura 2

En los sistemas precoordinados los términos están unidos de acuerdo con unas reglas sintácticas establecidas por el propio lenguaje. El prototipo de lenguaje precoordinado son las clasificaciones, nacidas de la necesidad de organizar los conocimientos según un orden establecido.

Los postcoordinados carecen de sintaxis; salvo excepciones, su vocabulario consiste en términos simples o unitérminos que se pueden conectar en el momento de la recuperación por medio de una sintaxis booleana. Atendiendo a los procedimientos seguidos para relacionar los conceptos al recuperar la información, podemos concluir que en los lenguajes precoordinados la relación entre los términos es gramatical, mientras que en un lenguaje postcoordinado la relación es lógica.

Los lenguajes precoordinados contribuyen mejor que los postcoordinados a conseguir precisión en la búsqueda, es decir, tienen mayor capacidad para rechazar los documentos irrelevantes en el momento de la recuperación porque las relaciones que se crean mediante el álgebra de Boole son genéricas y no impiden las falsas combinaciones. Como contrapartida, los postcoordinados tienen la ventaja de dar respuesta a necesidades de indización reales.

En cuanto a la **estructura**, se reconoce la existencia de dos tipos: *jerárquica* y *combinatoria*. A la estructura jerárquica responden determinadas clasificaciones, a la combinatoria, los léxicos documentales y los tesauros, si bien es cierto que en la organización multidimensional de estos últimos participa también la estructura jerárquica -representada por los índices sistematicos-. Presentados de forma esquemática, los lenguajes documentales, en función de su estructura, pueden ser:

<b>Estructura</b>	<b>Jerárquica</b>	Clasificaciones jerárquicas
	<b>Combinatoria</b>	Léxicos documentales
		Tesauros.

Figura 3

### 2.5. Concepto de tesoro documental

Atendiendo a la tipificación realizada en el anterior epígrafe, y más concretamente al criterio estructural, se puede considerar al tesoro como un lenguaje combinatorio, dado que consta de listas de términos que representan el ámbito científico o técnico a que se indica y posee además una serie de relaciones semánticas entre los términos que dan al lenguaje gran capacidad para la recuperación documental. Pero el tesoro participa también de la estructura jerárquica, en tanto en cuanto incluye en su composición un índice sistemático, que agrupa por áreas temáticas la información que contiene.

En virtud de la primera de las características estructurales que hemos citado, el tesoro tiene una serie de ventajas, entre las que podríamos destacar, la flexibilidad, la

capacidad de especialización, la multiplicidad de combinaciones que permite establecer entre los términos de su vocabulario, el alto nivel de control terminológico y la facilidad de revisión y puesta al día.

A nuestro juicio, podemos definir al tesauro como aquel *lenguaje documental de estructura jerárquico-combinatoria, de carácter especializado, que se basa en expresiones conceptuales llamadas descriptores, provistas de sus relaciones semánticas*. Se trata, pues, de un vocabulario controlado y estructurado al que se llega mediante la selección de términos del lenguaje natural. Está formado por listas de palabras -descriptores- estructuradas de forma que unas se relacionan con otras que también están en el tesauro. De forma precisa, Emilia Currás lo define en los siguientes términos: *lenguaje especializado, normalizado, post-coordinado, usado con fines documentarios, donde los elementos lingüísticos que lo componen se hallan relacionados entre sí sintácticamente y semánticamente*.<sup>5</sup>

La norma internacional ISO 2788-1986 *Directrices para el establecimiento y desarrollo de tesauros monolingües*, lo conceptualiza desde los puntos de vista de su función y de su estructura. Según su función, la norma entiende por tesauro *un instrumento de control terminológico que traduce a un lenguaje sistémico o documental el lenguaje natural empleado en los documentos y por los usuarios*. En razón de su estructura, *el tesauro consiste en un vocabulario controlado y dinámico de términos relacionados semántica y jerárquicamente, que se aplica a un campo específico del conocimiento*.

---

<sup>5</sup>CURRÁS, E. *Thesauros: lenguajes terminológicos*. Madrid: Paraninfo, 1991, p. 93.

A diferencia de un diccionario, un tesoro no define los términos que lo componen *salvo* en casos excepcionales para aclarar ambigüedades. El diccionario tiene por objeto definir los términos de la lengua, el tesoro, servir de puente entre el vocabulario utilizado por el indizador de los documentos y el usuario de los mismos.

Los tesoros documentales no son diccionarios exhaustivos. No aportan todos los términos relativos a un dominio, sino aquellos que podrían figurar en las consultas de los usuarios. El tesoro documental es por tanto un lenguaje documental, de indización y de búsqueda, concebido en función de las necesidades de información de los usuarios.

### **2.5.1. Funciones del tesoro**

El tesoro realiza tres funciones básicas a lo largo de las operaciones de tratamiento de la documentación: la función de normalización del vocabulario, la de inducción y la de representación.

La primera función tiene como fin unificar todo el vocabulario del tesoro para controlar todas las posibles entradas. Se puede conseguir mediante el control de formas, es decir, controlando los accidentes que puede generar el vocabulario y que son sinonimia y polisemia. Hay que controlar igualmente género y número de las expresiones. La inducción sirve para que el tesoro indique constantemente todas las alternativas posibles para entrar a recuperar: genéricas, específicas y asociadas. El tesoro induce, en efecto, al usuario hacia la consulta de temas en los que previamente no había pensado, gracias al establecimiento de una serie de referencias cruzadas que indican las relaciones asociativas, jerárquicas y preferenciales que se puedan dar entre la terminología que lo compone.

A propósito de esta función escriben Salton y McGill que *un thesaurus puede ampliar el vocabulario aumentando el nivel de exhaustividad en la recuperación. Con las relaciones jerárquicas proporcionadas por un tesoro en forma de términos más genéricos o más específicos, el vocabulario de indización y, por tanto, el de recuperación, pueden expandirse en varias direcciones añadiendo esos términos más amplios o más específicos o términos relacionados, en ocasiones*<sup>6</sup>.

Para que sugiera conceptos dentro del sistema, el vocabulario se estructura semánticamente, lo que se consigue mediante el establecimiento de relaciones paradigmáticas (verticales) y sintagmáticas (horizontales). Las paradigmáticas son relaciones invisibles que establece quien construye el tesoro, entre los distintos conceptos; son hilo conductor de la relación entre un término y otro. Las relaciones paradigmáticas deben existir entre todos los términos.

La función de representación, por último, la realiza al igual que los demás lenguajes documentales en virtud de la correspondencia que establece entre los términos que componen sus índices y los conceptos presentes en los documentos.

La figura 4 ilustra la labor que desempeña este lenguaje documental en las fases de indización y recuperación de la información en un sistema documental.

---

<sup>6</sup>SALTON, G. 6 MCGILL, M.J. **Introduction to modern information retrieval**. New York, etc.: McGraw Hill, 1983, p. 81.

EL PAPEL DEL TESAURO EN UN SISTEMA DE DOCUMENTACION

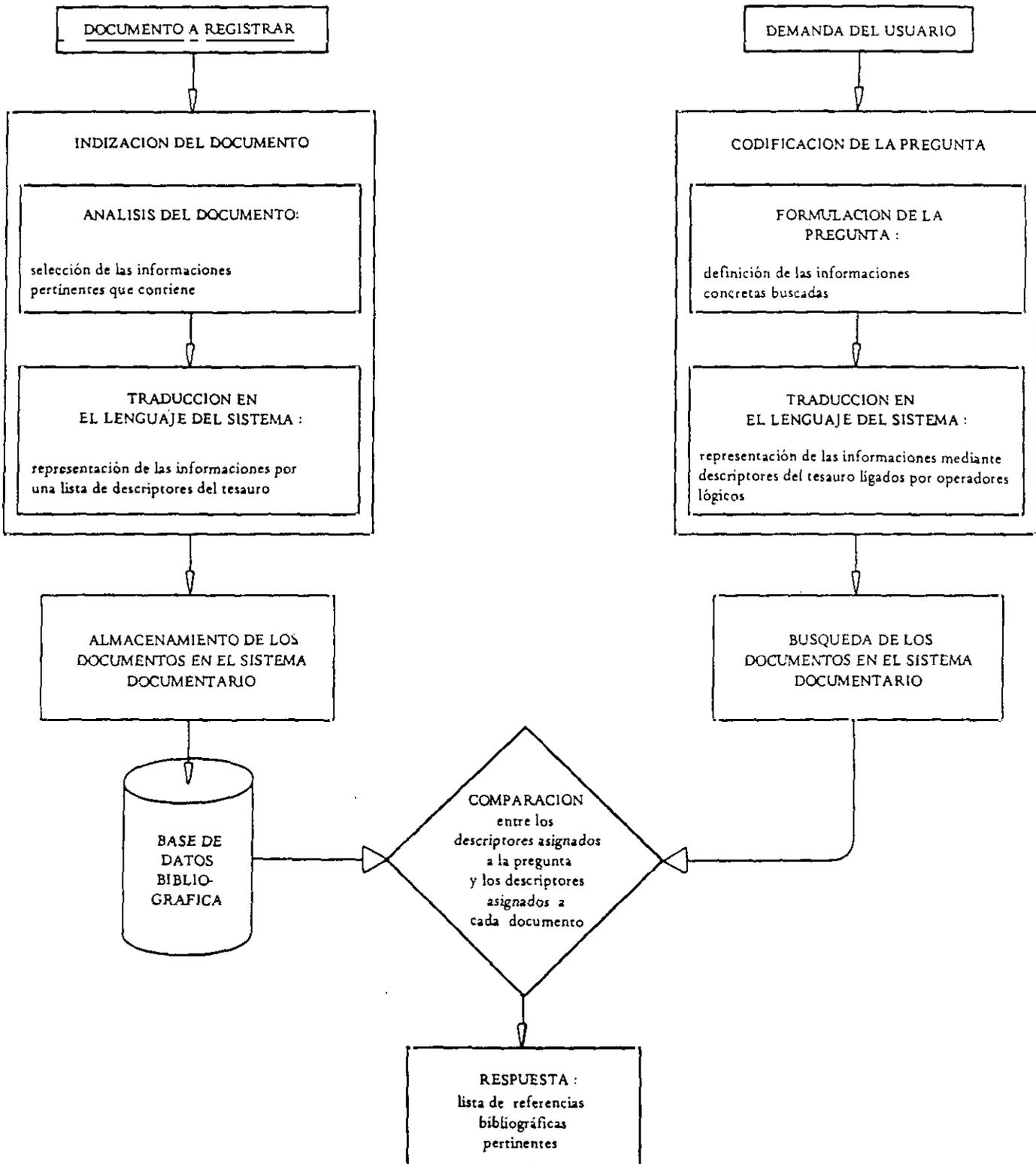


Figura 4. El papel del tesoro en un sistema de documentación (Tesauro Spines)

### 2.5.2. Composición del tesoro

Un tesoro se compone de una base léxica estructurada en relaciones jerárquicas, asociativas y de equivalencia, que observamos en el siguiente cuadro (figura 6) y que examinaremos detenidamente.

<b>Composición del tesoro</b>	Unidades léxicas	descriptores
		términos equivalentes
		palabras herramienta
	Relaciones semánticas	relaciones de equivalencia
		relaciones jerárquicas
		relaciones asociativas

Figura 5

Analizamos a continuación las unidades léxicas del tesoro, las relaciones semánticas se estudiarán en el epígrafe destinado a la metodología de elaboración de este lenguaje.

Un tesoro contiene, por lo general, algunos miles de descriptores y de varios cientos a varios miles de no descriptores. En total puede tener hasta 20.000 expresiones. El término medio es de unas 3.000.

La *Asociación Francesa para la Normalización (AFNOR) (Norma NZ 47-100 diciembre 1981)* define al **descriptor** como *Palabra o grupo de palabras incluidas en un*

*tesauro y escogidas de entre un conjunto de términos equivalentes para representar sin ambigüedad una noción contenida en un documento o en una petición de búsqueda documental.* Ej.: AGUAS SUBTERRÁNEAS

El **no descriptor** es llamado también término equivalente y término no preferente. Es sinónimo o cuasisinónimo del descriptor. No puede ser utilizado para indizar documentos ni para formular consultas, pero cada no descriptor reenvía a uno o a dos descriptores, por lo que aumentan la tasa de precisión, es decir, la pertinencia de los documentos que responden a una consulta.

Los no descriptores no constituyen en el tesauro sino puntos de acceso que facilitan el paso del lenguaje natural al lenguaje del sistema, permitiendo la elección de los descriptores pertinentes. Ej.:

Aguas freáticas

USE AGUAS SUBTERRÁNEAS

**Palabras herramienta:** Son términos que no tienen significado preciso si los empleamos solos, por lo que han de ser empleados con un descriptor, al menos. Ej.:  
PRIVATIZACIÓN

## Tipología del descriptor

<b>Tipología del descriptor</b>	Por su carga informativa	primarios secundarios infraconceptos
	Por su cobertura o contenido	onomásticos geográficos materias cronológicos
	Por su composición	simples compuestos

Figura 6

### 1. Por la **carga informativa**:

- *Primarios*: Términos o conjunto de términos que representan un concepto de manera unívoca (no hay duda de lo que queremos decir al emplearlo). Son significativos, relevantes, no hay posibilidad de ambigüedad, pudiendo aparecer aislados sin necesidad de aclaración. Ej.: ÁRBOL, CLIMA

- *Secundarios*: Van acompañando a otros descriptores. Parecen menos importantes que los primarios, pero son fundamentales en la estructura del lenguaje porque sirven para restringir la demanda. Ej.: PRIVATIZACIÓN DE LA INFORMACIÓN

- *Infraconceptos*: Son elementos exentos de significación pero, acompañados de un descriptor primario o secundario, pueden modificar a éste parcial o completamente. Son infraconceptos, por ejemplo, los prefijos o los sufijos. Ej.: HIPER, MEGA

## 2. Por la cobertura temática:

- *Onomásticos*: Representan un nombre de persona, de institución (corporativos), etc. Son fáciles de detectar y de recuperar en un texto determinado. Ej.: CRUZ ROJA

El uso de nombres propios permite una rápida y específica selección de los documentos, pero en muchos casos, un documento indizado por un nombre propio puede representarse o recuperarse por medio de descriptores más amplios, siempre que estuvieran incluidos en el lenguaje documental y establecidas las relaciones de generalidad y especificidad teniendo en cuenta dichos términos en relación con los demás contenidos en dicho lenguaje. Ej.: CERVANTES, podría representarse por LITERATURA ESPAÑOLA o por ESCRITORES.

- *Geográficos o territoriales*: abarcan todo tipo de conceptos vinculados con lugares y sitios (países, mares, montañas...). Al leer un texto saltan a la vista, pero hay que seleccionar entre los coyunturales y los no relevantes, al igual que en el caso de los onomásticos. Ej.: MAR ROJO

- *Temáticos o de materias*: Son los más difíciles de controlar y valorar y, a la vez, los más importantes. Representan cualquier contenido disciplinar. Ej.: FÍSICA

- *Cronológicos o temporales*: Representan fechas, períodos, etc. Ej.: SIGLO DE LAS LUCES

### 3. Por su **composición**:

- *Unitérminos o simples*: Representan conceptos mediante una sola palabra. Por su carácter unívoco, son idóneos para los lenguajes postcoordinados. Ej.: PLANTAS
  
- *Sintagmáticos o compuestos*: Representan un concepto utilizando un sintagma nominal o preposicional (conjunción de términos unidos por nexos). Se utilizan para eliminar la ambigüedad característica de los descriptores unitérminos. Ej.: PLANTAS MEDICINALES

Cada vez es más común el uso de descriptores sintagmáticos, es decir, términos precoordinados, ya que para la expresión de un concepto o la representación de una noción puede ser necesario combinar varias palabras o términos. Estos pueden ser combinados en el propio lenguaje del sistema y se hallarán, por tanto, precoordinados en el tesoro. O pueden ser combinados o postcoordinados durante la indización o la búsqueda documental.

Si cada una de las palabras que forman un descriptor compuesto están en el tesoro como descriptor simple, se le llama descriptor precoordinado. Si no existe un descriptor precoordinado que represente la noción deseada, el indizador tendrá que combinar varios descriptores para caracterizar el documento. Ahora bien, antes de combinar descriptores unitérminos para expresar un concepto, consultará sistemáticamente el tesoro para comprobar si ya existe un descriptor precoordinado que lo represente.

El grado de precoordinación influye decisivamente en la naturaleza de las jerarquías que se establezcan en el tesoro. Su campo de aplicación es decisivo: en tesauros especializados la ambigüedad terminológica es menor, se necesita menor nivel

de precoordinación dado que, en un contexto limitado, es más fácil el manejo de los términos que cuando se trata de temática diversa.

En relación con la utilización de términos simples o compuestos, la norma ISO determina en su epígrafe 7 que, por regla general, los descriptores deben representar en la medida de lo posible términos simples. Las nociones complejas pueden ser descompuestas en términos simples si alguno de los elementos de un término compuesto puede ser aceptado como descriptor. Ej.: ANÁLISIS DE DATOS (análisis + datos)

La misma norma señala que deben mantenerse los términos compuestos como descriptores:

a) Si la descomposición del término compuesto puede generar problemas de comprensión, por ser conocido en su forma compuesta. Ej.: CAJA NEGRA

b) Si la descomposición puede generar confusión o ambigüedad, y para evitar la amplitud semántica. Ej.: TRIÁNGULO DE LAS BERMUDAS

Sobre este asunto Unesco aconseja la utilización de la forma sustantiva si se trata de unitérminos y, en el caso de tratarse de descriptores sintagmáticos, al menos la base será un sustantivo. Añade que se han de evitar en lo posible los lazos gramaticales, artículos y preposiciones entre los descriptores compuestos; en español, para este objetivo, se puede recurrir a la adjetivación. Ej.:

**MINAS DE CARBÓN > MINAS CARBONÍFERAS**

Los descriptores sintagmáticos pueden formarse por unión morfológica o semántica:

1. Unión morfológica:

Se fusionan dos o más términos mediante preposiciones, artículos, sustantivos o adjetivación de uno de ellos, de manera que la estructura significativa no varía sustancialmente. El resultado debe ser identificable con respecto a los descriptores de los que partimos, es decir, varía el contenido pero permanece la forma. Ej.:

ENERGÍA + VIENTO > ENERGÍA EÓLICA

2. Unión lexicológica o semántica:

Se unen los términos por absorción de contenidos. Con este procedimiento la forma cambia parcial o totalmente. El resultado de la unión lexicológica será siempre un descriptor simple o unitérmino parcial o totalmente irreconocible en relación con los términos de los que se parte. Ej.:

PSICOLOGÍA + ESCRITURA > GRAFOLOGÍA

**Accidentes lingüísticos: polisemia y sinonimia:**

Al construir un tesoro es preciso fijar una terminología inequívoca, que supere sinonimias y polisemias, eligiendo, en la medida de lo posible, la transcripción más comúnmente utilizada.

La *polisemia*: Multiplicidad de acepciones de una palabra. La existencia de significantes con varios significados es perjudicial en el tesoro y debe ser controlada. Hay cuatro recursos para controlar o explicar la polisemia:

1. Introducción de un calificador, que concrete su significado, al lado del descriptor, entre paréntesis. Ej.: ROSA (color)

2. Asociación del término polisémico a otro descriptor del tesoro que ayude a precisar su significado, es decir, a contextualizarlo. La relación entre ambos términos se expresa mediante el símbolo TR (término relacionado). Ej.: ROSA TR BOTÁNICA

3. Sintagmatización o transformación del término descriptor en sintagma nominal. Ej.: ROSA SILVESTRE

4. A través de la utilización de un sinónimo que sustituya al término polisémico. Ej.: Megalópolis Use ÁREAS METROPOLITANAS

La *sinonimia* se produce cuando un solo concepto está representado por varios significantes. En un lenguaje documental la sinonimia es beneficiosa siempre que esté controlada, pues ofrece todas las entradas posibles al sistema documental. Lo más habitual es elegir una expresión como descriptor y mantener los sinónimos como no descriptores.

Según Chaumier<sup>7</sup>, en un tesoro pueden producirse dos tipos de sinonimia: lingüística y documental.

a) La sinonimia lingüística puede presentarse en estos casos:

1. Por variante ortográfica. Ej.: MÉXICO / MÉJICO
2. Por transliteración (distintos sistemas de conectar los diferentes alfabetos). Ej.: TCHEKOV, CEKOV, TCHEKOF
3. Sinonimia semántica. Ej.: WOLFRAMIO / TUNGSTENO
4. Equivalencia entre forma de expresión antigua y moderna. Ej.: ALMOFALLA / ALFOMBRA
5. Traducciones y préstamos. Ej.: SOFTWARE / PROGRAMA
6. Equivalencia entre el lenguaje popular y el científico. Ej.: OCULISTA / OFTALMÓLOGO
7. Variante de escritura. Ej.: C.E.E. / CEE

b) Sinónimos documentales: Designan términos de significado próximo, no considerados como equivalentes en el lenguaje corriente.

---

<sup>7</sup>CHAUMIER, J. *Análisis y lenguajes documentales: El tratamiento lingüístico de la información documental*. Barcelona: Mitre, 1986., pp. 82-84.

Denominados también cuasisinónimos, reagrupan bajo un descriptor varios términos por vecindad o inclusión pero no por auténtica sinonimia. Se aplica este procedimiento en los siguientes casos:

1. Si un término se considera demasiado específico para el tesoro, se sustituye por uno más general. Ej.: MARGARITAS / TULIPANES / ROSAS, se sustituyen por FLORES.

2. Términos variantes de un mismo concepto, se sustituyen por su sema principal. Ej.: PERIÓDICO / PERIODISTA por PERIODISMO

3. Términos antónimos: los transformamos en sinónimos uno del otro. Ej.: PACIENCIA por IMPACIENCIA

Siempre que se proceda de este modo, hay que hacer una nota aclaratoria del porqué.

### **2.5.3. Morfología de los descriptores**

Hacemos referencia a continuación a algunos de los epígrafes de la norma ISO 2788-1986, que consideramos esenciales en lo que a morfología de los descriptores se refiere.

**Epígrafe 4.4.: Los descriptores deben expresarse en mayúsculas y los no descriptores en minúsculas.**

### Epígrafe 6.2.1.: **Sustantivos y frases sustantivadas**

Bajo locuciones nominales la norma entiende los términos compuestos: locuciones adjetivas (plantas silvestres) o locuciones preposicionales (cuentos de hadas).

Los nombres o adjetivos modificadores pueden ser fuente de otros descriptores. Los tesauros deben en este caso mostrar las relaciones recíprocas entre el descriptor derivado y el término compuesto. Ej.:

FLORA ALPINA *TR* ALPES  
ALPES *TR* FLORA ALPINA

### Epígrafe 6.2.2.: **Adjetivos**

El empleo de adjetivos aislados debe ser evitado. Los adjetivos de locuciones adjetivas pueden dar lugar a reenvíos. Ej.:

CRIPTÓGAMAS (véanse las diferentes plantas criptógamas)

### Epígrafe 6.2.3.: **Adverbios**

Adverbios tales como *muy* o *grandemente* no son admitidos, salvo si forman parte de expresiones técnicas. Ej.: GRAN PRIMER PLANO

#### Epígrafe 6.2.4.: **Verbos**

El infinitivo de los verbos no debe ser empleado solo como término de indización. Las actividades deben ser sustantivadas. Ej.: DESTILACIÓN (no destilar).

#### Epígrafe 6.2.5.: **Abreviaturas y acrónimos**

Es preferible utilizar la forma desarrollada. Las siglas y acrónimos deben evitarse, salvo si se han sustantivado (LASER, RADAR) o si se conocen en cualquier idioma (ONU).

#### Epígrafe 6.3.: **Singular y plural**

Por regla general, los términos expresivos de entidades que pueden ser contadas pueden ir en plural. Ej.: BOSQUES

Los términos que representan entidades no contables deben ser expresados en singular. Ej.: AZÚCAR

También se expresan en singular los nombres de nociones abstractas. Ej.: IDEALISMO

Cuando un término designa una operación y el producto de la misma, se utiliza el singular para la operación y el plural para el producto, calificando además el primero con una expresión entre paréntesis. Ej.:

ALEACIÓN (proceso) / ALEACIONES

plantas o árboles en singular, su fruto en plural

MADROÑO (árbol) / MADROÑOS (frutos)

#### Epígrafe 7.4: **Orden natural de las palabras**

Las locuciones nominales, adjetivas y preposicionales deben ser presentadas en el orden de la lengua natural y no en la forma invertida: Ej.: *Ley del Talión* y no, *Talión, ley del*. La forma invertida puede dar lugar a un reenvío hacia la forma directa. Ej.:

TALIÓN, LEY DEL

*Use* LEY DEL TALIÓN

#### **2.5.4. Relaciones entre los términos del tesoro**

Una vez analizados los aspectos morfológicos que afectan a los descriptores, vamos a tratar en este epígrafe de las posibles relaciones que se pueden establecer entre los mismos, en el corpus del tesoro.

En un lenguaje documental los conceptos nunca se usan de forma aislada y su terminología se organiza en sistemas de conceptos pertenecientes a campos conceptuales. Dichos sistemas de conceptos facilitan la estructuración de la terminología propia de un campo del conocimiento e, incluso, facilitan la creación de equivalencias interlingüísticas.

El establecimiento de relaciones semánticas entre conceptos es el paso previo a la creación de sistemas de conceptos para la confección de cualquier lenguaje documental.

En un lenguaje documental y, por tanto, en un tesoro, las relaciones entre conceptos se dividen en dos grandes grupos, las relaciones concepto-concepto y las relaciones concepto-término dando origen, a su vez, a tres tipos de relaciones semánticas: relaciones jerárquicas y asociativas, que responden a la categoría relacional concepto-concepto; y relaciones de equivalencia, establecidas entre concepto-término.

Las relaciones basadas en la secuencia de conceptos subordinados, es decir, las relaciones jerárquicas, se desarrollan en el eje vertical, o paradigmático, mientras que las asociativas se fundamentan en la secuencia de conceptos coordinados, ésto es, en el eje horizontal, porque se establecen entre conceptos del mismo nivel jerárquico.

#### **2.5.4.1. Relaciones jerárquicas o paradigmáticas**

Las relaciones jerárquicas sirven para expresar las relaciones de superioridad e inferioridad entre los conceptos. En un tesoro estas relaciones permiten reunir aquellos conceptos -descriptores- que unos con respecto de otros son más generales o más específicos, colocándolos en niveles superiores o inferiores y formando así un conjunto jerárquico preciso. Pueden ser de amplitud y especificidad y partitivas. En el primer caso, la relación se establece entre un concepto que se considera genérico y aquellos que, siendo coordinados entre si, son subordinados del genérico. Ej.:

## ENERGÍAS

Energías convencionales

Energías alternativas

Por tratarse de una relación transitiva, no simétrica, se puede incorporar fácilmente en los sistemas de recuperación de información. El establecimiento de jerarquías facilita, en efecto, en el momento de la indización, la elección de los términos más específicos que definan un concepto, eliminando la posibilidad de ruidos en la búsqueda.

Una relación genérica propiamente dicha es la monojerárquica, en la cual un concepto se encadena jerárquicamente sólo a otro. Ej.: Si todas las energías alternativas son ENERGÍAS y si algunas ENERGÍAS son energías alternativas, el concepto energías alternativas es específico del concepto ENERGÍAS.

Pueden darse casos de polijerarquía o multijerarquía cuando un mismo concepto pertenece a más de una jerarquía. Por ejemplo: El nervio óptico puede incluirse en la categoría semántica relativa al SISTEMA NERVIOSO, y puede pertenecer jerárquicamente al campo destinado a la ÓPTICA.

No todos los especialistas coinciden en la necesidad y conveniencia de incluir relaciones polijerárquicas en la estructura de un tesoro. Lancaster, por ejemplo desaconseja su empleo<sup>8</sup>, mientras que Soergel<sup>9</sup> considera la polijerarquía un medio de

---

<sup>8</sup>LANCASTER, F.W. **Thesaurus construction and use: a condensed course**. Paris: Unesco, 1985, p. 41.

<sup>9</sup>SOERGEL, D. **Indexing languages and thesauri: Construction and maintenance**. Los Angeles: Melville, 1974, p. 254.

representación adecuado de los conceptos, capaz de eliminar problemas terminológicos.

Por lo que se refiere a las relaciones jerárquicas partitivas, el descriptor inferior es una parte particular del descriptor superior. Ej.:

EUROPA y ESPAÑA  
FLOR y PÉTALOS

En un tesoro puede llegar a haber hasta 15 niveles jerárquicos. El orden en el que se colocan los descriptores del mismo nivel jerárquico dependientes de un mismo término, es el siguiente: Orden alfabético, el que utilizan la mayoría de los tesoros. Ej.:

METALES PRECIOSOS  
TE ORO  
PLATA  
PLATINO

Van Slype<sup>10</sup> considera aplicable el orden alfabético cuando no se pueda utilizar ningún otro modo de ordenación, y propone los siguientes criterios alternativos:

- *orden cronológico, aplicable*

- *en materia histórica y estilística*
- *en procesos que se desencadenan secuencialmente (ej.: producción, distribución, consumo).*

---

<sup>10</sup>SLYPE, G. v. Op. cit., pp. 53-54.

- *orden de evolución, aplicable especialmente en biología;*
- *orden de complejidad creciente, aplicable, por ejemplo, en el dominio de las ciencias (geometría euclídea, geometría no euclídea...);*
- *orden de dimensión creciente (por ejemplo: composición musical: solo dúo, trío, cuarteto...) o decreciente (por ejemplo: subdivisión territorial, país, región, provincia, municipio...);*
- *ordenación en el espacio (por ejemplo, disposición de los países en orden de contigüidad);*
- *orden de preferencia o de frecuencia de uso en los documentos tratados por el servicio (por ejemplo: religión: catolicismo, protestantismo, judaísmo, islamismo, budismo...);*
- *orden tradicional (por ejemplo, ordenación de las ciencias: física, química, biología...);*
- *y, por último, el orden alfabético, aplicable cuando no se pueda utilizar ningún otro modo de ordenación.*

#### **2.5.4.2. Relaciones asociativas o sintagmáticas**

Las relaciones asociativas o a posteriori, también se denominan horizontales porque se establecen entre conceptos del mismo nivel jerárquico. Relacionan descriptores que no suponen sinonimia ni jerarquía y se trata, pues, de una relación de afinidad semántica. Ambos no se incluyen ni son sinónimos, pero por asociación de ideas envían a otros conceptos.

Es el tipo de relación más difícil de definir porque, al no existir una norma precisa para su establecimiento, el criterio para su selección es fundamentalmente pragmático.

Un descriptor puede tener una, ninguna, dos o más relaciones asociativas.

## RECURSOS HÍDRICOS

TR AFOROS

TR AGUA

TR HIDROLOGÍA

Las relaciones de equivalencia y las relaciones jerárquicas sólo se establecen entre términos del mismo tipo. No es posible asociar, por ejemplo, un término genérico y un término específico, es decir, no se crearán relaciones asociativas entre descriptores de la misma cadena jerárquica. Las relaciones asociativas, de reciprocidad directa, no pueden conectar descriptores y no descriptores, en cambio, pueden enlazar determinado descriptor con otro cualquiera.

Si un término se asocia con otro de nivel jerárquico superior, se supone que está relacionado también con todos los que están subordinados a él.

Aitchison y Gilchrist<sup>11</sup>, proponen casos orientativos en los que se pueden establecer relaciones de asociación:

1. Entre conceptos que representan disciplinas y los objetos estudiados por esa disciplina. Ej.: SISMOLOGÍA-TERREMOTO

2. Entre operaciones y procesos y sus agentes o instrumentos. Ej.: VELOCIDAD-VELOCÍMETRO

---

<sup>11</sup>AITCHISON & GILCHRIST. *Thesaurus construction, a practical manual*. London: ASLIB, 1987, pp. 44-49.

3. Ocupaciones y personas. Ej.: DEPORTE-DEPORTISTA
  
4. Acciones y productos o efectos de las acciones. Ej.:  
INVESTIGACIÓN-DESCUBRIMIENTOS
  
5. Una acción y su objeto paciente. Ej.:  
ANÁLISIS DOCUMENTAL-DOCUMENTO
  
6. Partes del todo con partes de otro todo. Ej.: VOLANTE-MANOS
  
7. Objetos relacionados con sus propiedades. Ej.: HIELO-FRIALDAD
  
8. Objetos relacionados con sus orígenes. Ej.: AGUA-MANANTIAL
  
9. Efectos y causas. Ej.: EROSIÓN-VIENTO
  
10. Cosas y/o acciones y sus antiagentes. Ej.: LLUVIA-PARAGUAS
  
11. Materias primas y sus productos. Ej.: MADERA-MUEBLE
  
12. Acciones y propiedades asociadas a esas acciones. Ej.:  
INFORMACIÓN-RECUPERACIÓN DE LA INFORMACIÓN
  
13. Conceptos antónimos. Ej.: ESTABILIDAD-INESTABILIDAD

Las recomendaciones de Unesco admiten también la relación entre conceptos que, perteneciendo a una misma cadena jerárquica, tienen una significación poco clara cuando se consideran aisladamente.

#### **2.5.4.3. Relaciones de equivalencia**

Agrupan términos preferenciales a través de la localización de sinónimos en la *terminología/vocabulario*, que no deben aparecer simultáneamente en el tesoro como descriptores autorizados.

Detectadas las equivalencias semánticas se elige uno de los términos como representativo del concepto, dándole la categoría de término preferente o descriptor. Los términos equivalentes se remiten al principal.

Capa de ozono

*Use* OZONOSFERA

OZONOSFERA

*UP* capa de ozono

#### **2.5.5. Formas de presentación de los descriptores**

Los términos que componen el tesoro y sus relaciones se pueden representar de varias formas, siendo las más comunes las siguientes:

1. Presentación alfabética
2. Presentación sistemática
3. *Presentación gráfica*

Describimos seguidamente los tres sistemas, si bien consideramos que, para que un tesoro constituya una precisa herramienta de recuperación de información, es menester que en su estructura estén presentes, al menos, dos de las citadas formas de presentación: alfabética y sistemática o alfabética y gráfica; de tal manera que, cuando la presentación de los términos se lleve a cabo sistemática o gráficamente, el tesoro ha de estar provisto de un índice alfabético que permita una primera toma de contacto con la información presente en el tesoro y remita, mediante un código, a la parte sistemática o gráfica.

La mayor ventaja que ofrece el orden sistemático es la ayuda que aporta a la búsqueda de información y a la investigación, pero requiere apoyarse en el orden alfabético que tiene la ventaja de resultarnos familiar.

Como complemento a estas formas de presentación es aconsejable elaborar un índice permutado general en donde estarán todos los descriptores alfabetizados tantas veces como términos los compongan. Es igualmente importante incluir, al comienzo de este lenguaje documental, una introducción en la que se explique cual es su intencionalidad, área temática cubierta y grado de especificidad de los conceptos incluidos, extensión, número de descriptores y no descriptores, los recursos para establecer las relaciones, método, estructura, fuentes utilizadas para construir el tesoro, indicando para cada fuente si todos o sólo parte de los términos contenidos en ella fueron incluidos en el tesoro, bibliografía utilizada, los códigos y su explicación, explicación igualmente de todas las normas de uso, nivel de compatibilidad con otros tesauros del área y con

macrotesauros. La introducción incluirá también su justificación e información sobre los procedimientos de actualización.

#### **2.5.5.1. Presentación alfabética**

El índice alfabético estructurado (figura 7) contiene descriptores y no descriptores ordenados de forma alfabética. Permite al indizador verificar la existencia de aquellos posibles descriptores que se le ocurran mientras que la temática le permitirá descubrir los descriptores pertinentes que no es capaz de memorizar. Con cada descriptor aparece el conjunto completo de sus relaciones semánticas incluida toda su jerarquía, es decir, todos los niveles de términos más genéricos y más específicos, numerados, para indicar su grado jerárquico con relación al descriptor de que se trate. Por lo tanto, cada entrada lleva:

- Indicación del campo semántico al que pertenece un descriptor
- Notas aclaratorias o de aplicación
- Equivalencias semánticas (relación de no descriptores o términos no preferentes)
- Los descriptores genéricos según uno, varios o todos los niveles de jerarquía ascendente
- Los descriptores específicos según uno, varios o todos los niveles de jerarquía descendente
- Los descriptores asociados

La indicación detallada de estas relaciones semánticas entre los términos caracteriza al tesoro y le diferencia de un léxico documental o de un simple glosario.

La incorporación de los términos sinónimos y cuasisinónimos de los descriptores hace que el índice alfabético sea un eficaz útil para la indización y recuperación, ya que, al incorporar dichas entradas, la probabilidad de encontrar el término adecuado es mucho mayor.

Las relaciones pueden ser representadas por los signos alfabéticos *USE*, *UP*, *TG*, *TE* y *TR* o igualmente por medio de símbolos, que permiten su identificación independientemente del idioma:

---	Relación de equivalencia	<i>USE (Use)</i>
=	" "	<i>UP (Usado por)</i>
<	Relación jerárquica	<i>TG (Término genérico)</i>
>	" "	<i>TE (Término específico)</i>
-	Relación asociativa	<i>TR (Término relacionado)</i>

En el índice alfabético estructurado los descriptores asociados y los no descriptores van precedidos por el operador o símbolo de relación correspondiente y dispuestos por orden alfabético, generalmente. Ej.:

## MÉTODOS EDUCATIVOS

*UP* Métodos pedagógicos

*TG* PLANIFICACIÓN EDUCATIVA

*TE* JUEGOS EDUCATIVOS

*TR* ENSEÑANZA

Los descriptores genéricos y específicos van precedidos por la sigla de relación correspondiente y dispuestos, dentro de cada nivel jerárquico, por orden alfabético; cada nivel jerárquico adicional está marcado por un sangrado hacia la derecha o por una

codificación numérica. Ej.:

ECOSISTEMAS ACUÁTICOS

TE1 HUMEDALES

TE2 ESTEROS

TE2 MARISMAS

TE2 PANTANOS

#### **2.5.5.2. Presentación sistemática**

El índice sistemático contiene los descriptores estructurados en función de categorías o jerarquías. La parte sistemática da una visión de conjunto de la información contenida en el tesoro y permite conocer en qué medida se ha detallado un tema. *Un thesaurus* -escriben Salton y McGill<sup>12</sup>- *propone una agrupación o clasificación de los términos utilizados en una área temática en categorías conocidas como temas, campos o clases.*

Con esta agrupación jerárquica de los descriptores se puede efectuar un control general rápido de los campos temáticos tratados y escoger el apropiado nivel de generalidad en el momento de indizar o al formular una ecuación de búsqueda.

Los descriptores se sitúan en sus áreas contextuales donde son jerarquizados por categoría semántica y dotados de distintos tipos de relaciones y reenvíos. Así es como después los encontraremos rodeados de sus conceptos más próximos. Para llevar a cabo esta parte, hay que ver posibles relaciones término a término y campo a campo.

---

<sup>12</sup>SALTON, G. & MCGILL, M.J. Op. cit., p. 76.

El establecimiento de categorías se puede realizar en base a facetas o a campos temáticos. Esta agrupación de los descriptores en facetas o temas es fundamental para poder manejar este lenguaje, que puede contar con varios miles de términos.

La presentación mediante campos temáticos organiza los conceptos según las relaciones género a especie, al igual que lo hacen los sistemas de clasificación jerárquicos. Los descriptores se presentan en esquemas arborescentes con ramificaciones sucesivas sin que, en ocasiones, exista un rigor lógico en su estructura, lo que se justifica en el seno de este lenguaje documental por razones pragmáticas.

Cuando la distribución de los términos del tesoro se realiza en base a facetas, se procede a la reagrupación en base a categorías artificiales que no se adaptan a la realidad cultural. La clasificación por facetas considera cada sujeto clasificable desde determinados puntos de vista, lo que posibilita la agrupación de temas por afinidades, formando conjuntos que tengan una característica común.

Ranganathan, creador del sistema de clasificación por facetas, identificó las siguientes: Personalidad, Materia, Energía, Espacio y Tiempo, capaces de realizar agrupaciones en campos semánticos. Las facetas utilizadas más comúnmente en la elaboración del área categorial de los tesauros son: fenómeno, proceso, materiales, organización, ser vivo, equipamiento, propiedad y disciplina. Dice Lancaster que *el tesoro por facetas presenta una ventaja sobre cualquier otro tipo de tesoro: puede utilizarse para la ordenación de libros en las estanterías de una biblioteca especializada, además de para la indización de documentos en una base de datos*<sup>13</sup>.

---

<sup>13</sup>LANCASTER, F. **El control del vocabulario en la recuperación de información**. Valencia: Universitat, 1995, p. 109.

Ya sea en base a temas o en base a facetas, la representación del área jerárquica del tesoro se puede hacer en la misma forma que el índice alfabético estructurado, salvo que el orden de los descriptores estará dispuesto de forma jerárquica y no alfabética (figura 8), o bien se representará gráficamente, tal y como aparece en las figuras 9, 10 y 11, correspondientes a los modelos de representación gráfica de uso más común y que estudiamos a continuación.

### **2.5.5.3. Presentación gráfica**

las presentaciones en listas sólo permiten examinar la estructura del campo semántico de forma lineal, por el contrario, las representaciones gráficas exponen la estructura semántica de cada campo semántico dentro de una hoja de papel y se sirven de las dos dimensiones de esta superficie. Las tres formas más usuales de representación gráfica son:

- a) estructura arborescente
- b) diagramas de flechas
- c) terminogramas

Los tesauros de representación gráfica se complementan con un índice alfabético que contiene notas aclaratorias, reenvíos entre términos preferentes y no preferentes y referencias a términos genéricos, específicos y relacionados; por ello, en este tipo de representación no es necesario asignar símbolos a cada término en particular. Así pues, los sistemas gráficos muestran sobre todo las relaciones jerárquicas existentes entre descriptores y por lo general, los reenvíos hacia términos asociados se indican en los márgenes del esquema gráfico.

- AMINOACIDOS  
DOPAMINA  
NORADRENALINA
- ACIDO RIBONUCLEICO 060101**  
UP ARN  
RNA  
TG ACIDOS NUCLEICOS  
TR AMINOACIDOS
- ACIDOS 0601**  
TE ACIDOS NUCLEICOS  
AMINOACIDOS  
TR ANALGESICOS  
LSD
- ACIDOS NUCLEICOS 060101**  
TG ACIDOS  
TE ACIDO DESOXIRIBONUCLEICO  
ACIDO RIBONUCLEICO  
ADENOSINA
- ACNE 120510**  
TG TRASTORNOS DERMATOLOGICOS  
TR ALERGIA
- ACOMODACION 080201**  
NA Proceso mediante el cual el sujeto ajusta sus esquemas y patrones de conducta a las condiciones externas, (Piaget).  
TG DESARROLLO COGNITIVO  
TR ASIMILACION  
DESARROLLO OPERATORIO  
DESARROLLO PERCEPTIVO  
EQUILIBRACION
- ACONTECIMIENTOS VITALES  
USE **SUCESOS VITALES**
- ACOSO SEXUAL  
USE **HOSTIGAMIENTO SEXUAL**
- ACROFOBIA 120103**  
NA Miedo intenso a las alturas.  
UP MIEDO A LAS ALTURAS  
TG FOBIAS  
TR TRASTORNOS MENTALES
- ACROMATISMO  
USE **CEGUERA PARA EL COLOR**
- ACTAS 170101**  
TG DOCUMENTOS PRIMARIOS
- ACTING OUT  
USE **ACTUACION**
- ACTITUD ANTE LA ENFERMEDAD 100201**  
TG ACTITUDES  
TR ACTITUD DEL PACIENTE  
CONDUCTA EN LA ENFERMEDAD  
ENFERMEDAD
- ACTITUD ANTE LA MUERTE 100201**  
TG ACTITUDES  
TR ACTITUD DEL PACIENTE  
CONDUCTA EN LA ENFERMEDAD  
EUTANASIA
- ACTITUD DE LOS PADRES 090301**  
UP ACTITUD PARENTAL  
TG RELACIONES FAMILIARES  
TE SOBREPOTECCION  
TR CRIANZA  
MADRES
- PADRES
- ACTITUD DE RESPUESTA 020101**  
NA Tendencia o estilo deliberado o inadvertido a responder a los items de un test de forma característica (p. ej. con respuestas socialmente aceptadas) que restan validez a los resultados.  
TG APLICACION DE TESTS  
TR ELECCION  
ESTILO COGNITIVO  
RESPUESTAS
- ACTITUD DEL CLIENTE  
USE **ACTITUD DEL PACIENTE**
- ACTITUD DEL ESTUDIANTE 150402**  
TR ABANDONO DE ESTUDIOS  
ADAPTACION ESCOLAR  
ALUMNOS  
ASISTENCIA A LA ESCUELA  
CARACTERISTICAS DE LOS ESTUDIANTES  
CONDUCTA DEL ESTUDIANTE  
ESTUDIANTES  
RELACION PROFESOR-ALUMNO
- ACTITUD DEL PACIENTE 130502**  
UP ACTITUD DEL CLIENTE  
TG PROCESO TERAPEUTICO  
TR ABANDONO DEL TRATAMIENTO  
ACTITUD ANTE LA ENFERMEDAD  
ACTITUD ANTE LA MUERTE  
CARACTERISTICAS DE LOS PACIENTES  
CONTRATRANSFERENCIA  
TRANSFERENCIA PSICOTERAPEUTICA
- ACTITUD DEL PROFESOR  
USE **CONDUCTA DEL PROFESOR**
- ACTITUD DEL PSICOTERAPEUTA  
USE **ACTITUD DEL TERAPEUTA**
- ACTITUD DEL TERAPEUTA 130502**  
UP ACTITUD DEL PSICOTERAPEUTA  
TG CARACTERISTICAS DE LOS TERAPEUTAS  
TR CONTRATO TERAPEUTICO  
CONTRATRANSFERENCIA  
PROCESO PSICOTERAPEUTICO  
RELACIONES PACIENTE-TERAPEUTA  
RELACIONES TERAPEUTA-FAMILIA  
TERAPEUTAS
- ACTITUD PARENTAL  
USE **ACTITUD DE LOS PADRES**
- ACTITUD SEXUAL 090501**  
NA Opiniones y creencias acerca de la conducta y desarrollo sexual.  
TG CONDUCTA SEXUAL  
TR ABSTINENCIA SEXUAL  
ATRACCION SEXUAL  
DESARROLLO PSICOSEXUAL  
DIFERENCIAS ENTRE LOS SEXOS  
DISCRIMINACION SEXUAL  
ESTEREOTIPOS SEXUALES  
IDENTIDAD SEXUAL  
MACHISMO
- ACTITUDES 100201**  
UP CREENCIAS  
TE ACTITUD ANTE LA ENFERMEDAD  
ACTITUD ANTE LA MUERTE  
ACTITUDES SOCIALES  
CAMBIO DE ACTITUD  
TR ACTITUDES POLITICAS  
ALTRUISMO  
ASPIRACION

Extracto del índice alfabético estructurado  
del Tesauro ISOC de Psicología

Figura 7

- ....RETROALIMENTACION
  - .....BIORRETROALIMENTACION
  - .....CONOCIMIENTO DEL RESULTADO
- ....SENSIBILIDAD
- ....UMBRALES
  - .....UMBRAL AUDITIVO
  - .....UMBRAL DE DOLOR
  - .....UMBRAL VISUAL

## 06 .FARMACOLOGIA Y PSICOFARMACOLOGIA

- 0601 ..ACIDOS
  - 060101 ...ACIDOS NUCLEICOS
    - ....ACIDO DESOXIRRIBONUCLEICO
    - ....ACIDO RIBONUCLEICO
    - ....ADENOSINA
  - 060102 ...AMINOACIDOS
    - ....TIROSINA
- 0602 ..ALCOHOLES
  - 060201 ...ETANOL
- 0603 ..AMINAS
  - 060301 ...AMITRIPTILINA
  - 060302 ...HISTAMINA
  - 060303 ...SEROTONINA
    - ....SULFO-ADENOSIL-L-METIONINA
- 0604 ..DROGAS
  - 060401 ...ALUCINOGENOS
    - ....CANNABIS
      - .....HACHIS
      - .....MARIHUANA
    - ....LSD
  - 060402 ...BEBIDAS ALCOHOLICAS
  - 060403 ...BEBIDAS NO ALCOHOLICAS
  - 060404 ...CAFEINA
  - 060405 ...COCAINA
  - 060406 ...INHALANTES
  - 060407 ...NARCOTICOS
    - ....OPIACEOS
      - .....CODEINA
      - .....HEROINA
  - 060408 ...TABACO
- 0605 ..ELEMENTOS QUIMICOS
  - 060501 ...IONES
  - 060502 ...LITIO
  - 060503 ...CALCIO
- 0606 ..ENZIMAS
- 0607 ..HORMONAS
  - 060701 ...HORMONAS SEXUALES
  - 060702 ...HORMONAS TIROIDEAS
- 0608 ..INHIBIDORES ENZIMATICOS
  - 060801 ...MAO
- 0609 ..MEDICAMENTOS
  - 060901 ...ANALGESICOS
    - ....METADONA
    - ....MORFINA
  - 060902 ...ANESTESICOS
  - 060903 ...ANOREXICOS
  - 060904 ...ANTIBIOTICOS
  - 060905 ...ANTICONVULSIVOS
  - 060906 ...ANTIESPASMODICOS
  - 060907 ...DIURETICOS
  - 060908 ...PSICOFARMACOS
    - ....DEPRESORES DEL SNC

Extracto del índice sistemático del  
Tesauro ISOC de Psicología

### **2.5.5.3.1. Diagramas arborescentes**

Es la forma de presentación gráfica más simple. Suelen construirse pocos diagramas por tesauro, como en el caso del Urbamet<sup>14</sup> (figura 9). En estos tesauros la estructura jerárquica sólo aparece en los diagramas, el listado alfabético se limita a darnos el término más amplio en el que se incluye el descriptor buscado para, en pasos sucesivos, acabar en los descriptores más genéricos que dan nombre a los distintos diagramas.

### **2.5.5.3.2. Diagramas de flechas (Figura 10)**

Se usa una hoja por campo semántico. En el interior del rectángulo figuran los descriptores que pertenecen a ese campo, así como sus relaciones jerárquicas y asociativas con descriptores situados dentro del mismo grupo.

Fuera del rectángulo se sitúan los descriptores que pertenecen a otros campos, pero que están ligados a los que figuran dentro del rectángulo por una relación jerárquica (en el caso de polijerarquía) o por una relación asociativa.

Se enlazan con los descriptores interiores por medio de flechas unidireccionales, en el caso de relaciones jerárquicas, o líneas rectas, cuando exista entre los términos una relación asociativa. Cada uno de estos descriptores lleva el código del campo semántico al que pertenece.

---

<sup>14</sup>THESAURUS URBAMET. 6ª ed. Paris: Ministère de l'Équipement des Transports et du Tourisme; Council Régional d'Île-de-France, 1993.

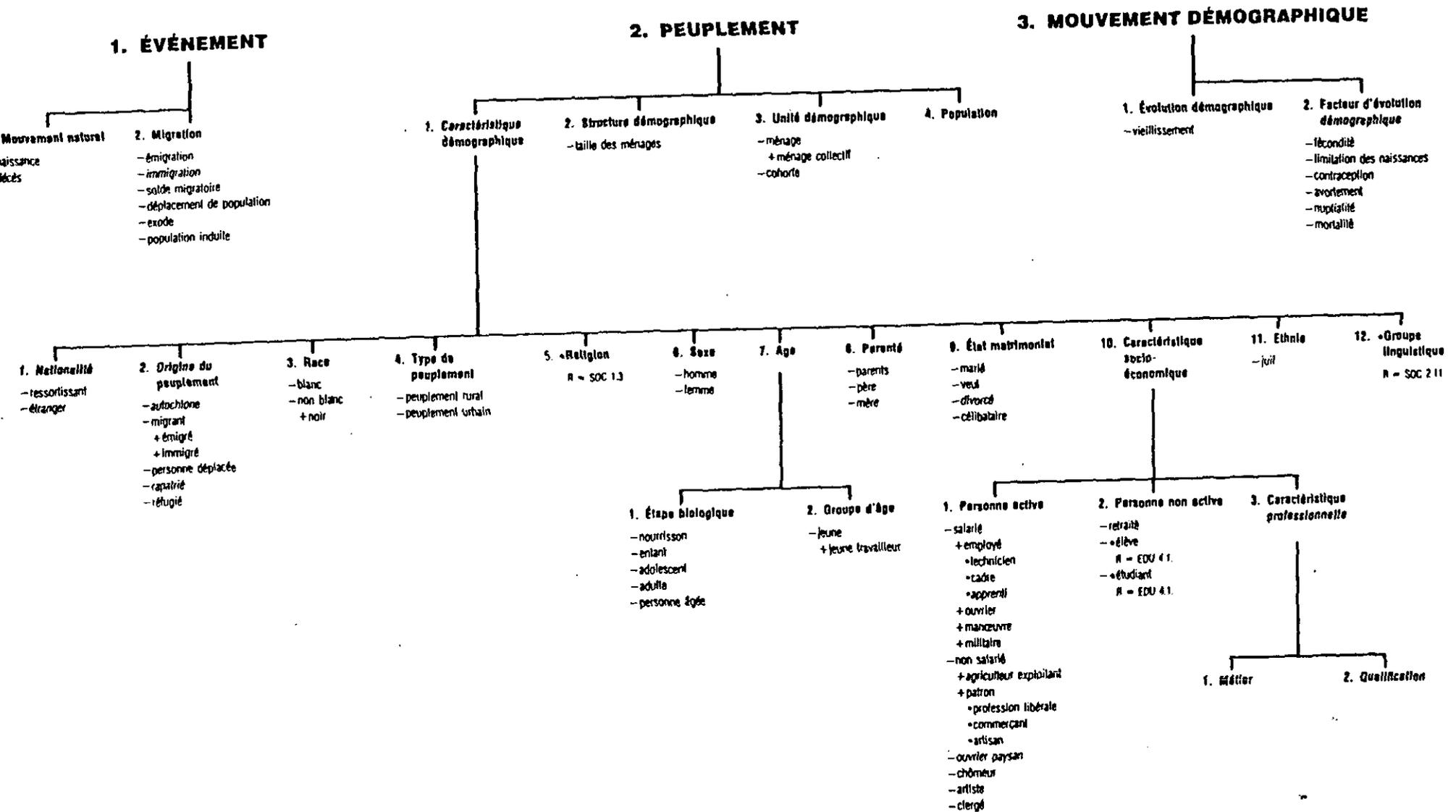


Figura 9. Diagrama arborescente. Tesauru URBAMET

**DÉMOGRAPHIE**

Figure 24. Graphic display  
Arrowgraph  
IRRD thesaurus

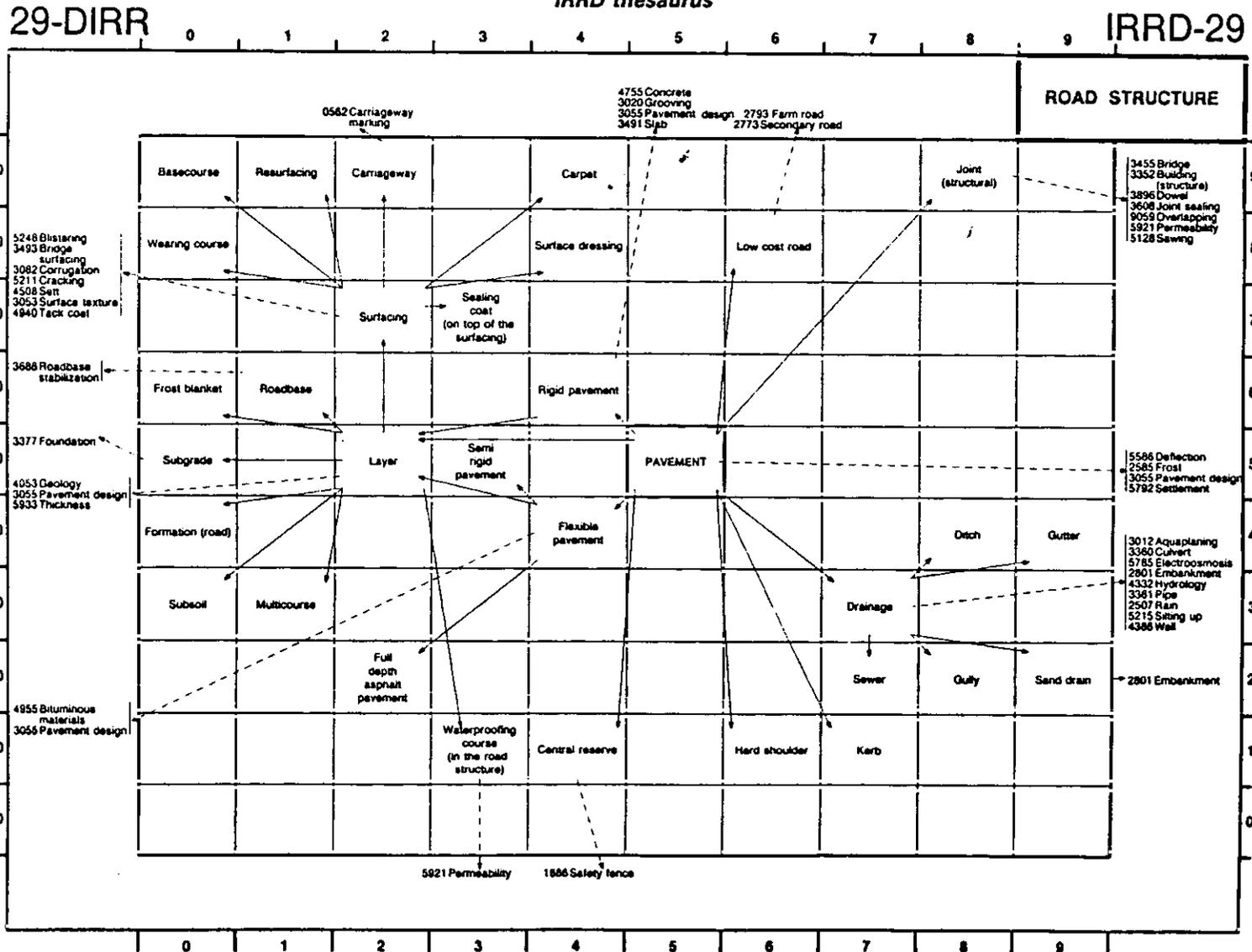


Figura 10. Diagrama flechado de descriptores. Thesaurus International Road Research Documentation (IRRD)

### 2.5.5.3.3. Terminogramas (Figura 11)

Reúnen los descriptores en grupos, cada uno de los cuales se corresponde con una disciplina o microdisciplina específicas: física, literatura, gramática...

En el caso de los terminogramas, en el interior del rectángulo hay un pequeño rectángulo - *cartucho* - por cada descriptor cabeza de jerarquía. Dentro se encuentran los descriptores en orden descendente de jerarquía: un sangrado a la derecha para cada nivel jerárquico y ordenación alfabética de los descriptores dentro de su propio nivel jerárquico.

Las relaciones asociativas, que sólo existen entre descriptores de cadenas jerárquicas distintas, se representan por líneas rectas entre descriptores asociados.

En el exterior del rectángulo están los descriptores de otros campos semánticos, con el número de código de su propio campo, enlazados con los descriptores del interior del rectángulo mediante líneas rectas o flechas.

En un terminograma se pueden incluir más o menos 80 descriptores y en un diagrama de flechas, unos 40. Es más práctico el primer procedimiento porque permite descomponer el tesoro en menos campos semánticos, lo que facilita la consulta.

#### 2.5.5.4. Índices auxiliares

Además de las formas de presentación mencionadas anteriormente, existe la posibilidad de incluir índices auxiliares en el tesoro, que facilitan el acceso a la información contenida en los índices alfabético y sistemático. En un tesoro es obligatorio al menos uno de estos índices, el permutado general, que contiene todas las entradas existentes en dicho lenguaje documental. El tesoro puede contener, igualmente, otros índices auxiliares: onomásticos, cronológicos o geográficos, compuestos por descriptores de esta tipología, que no tienen cabida en los diferentes campos temáticos que integran su estructura.

El índice rotativo o permutado (Figura 12) tiene como objetivo concentrar las palabras. Consiste en una simple relación alfabética de los descriptores para facilitar su búsqueda inicial y remitir al usuario, por medio de un código, a la parte sistemática.

Recibe la denominación de permutado porque los descriptores sintagmáticos se encuentran en este índice ordenados alfabéticamente tantas veces como términos los componen, de esta forma podemos encontrar los términos compuestos con la ayuda de los demás elementos que forman el sintagma, lo que el orden natural de las palabras que figuran en una simple lista alfabética no permite descubrir. Por ej.: cuencas fluviales se puede localizar por la [c] de cuencas y por la [ff] de fluviales. La estructura de este tipo de índices es la que diferencia a los tesoros de los léxicos.

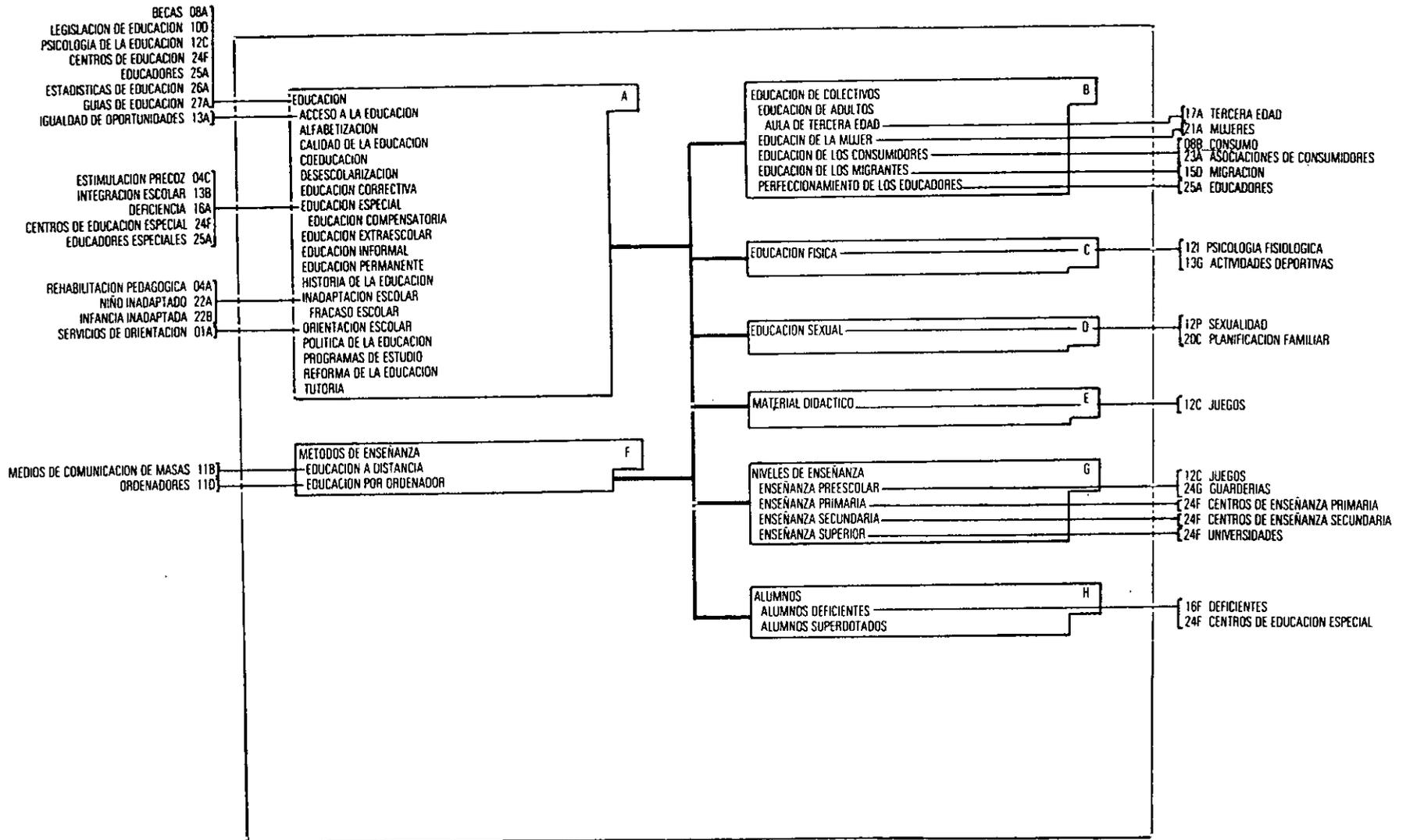


Figura 11. Terminograma de descriptores. Tesauro de Servicios sociales

1 CURSO DE EGB 150216  
 2 CURSO DE EGB 150216  
 TRISOMIA 21  
 3 CURSO DE EGB 150216  
 4 CURSO DE EGB 150216  
 5 CURSO DE EGB 150216  
 6 CURSO DE EGB 150216  
 7 CURSO DE EGB 150216  
 8 CURSO DE EGB 150216  
 NIÑOS ABANDONADOS 090301  
 ABANDONO DE ESTUDIOS 150401  
 ABANDONO DE HIJOS  
 ABANDONO DEL HOGAR 090301  
 ABANDONO DEL TRATAMIENTO 130505  
 ABDOMEN 050101  
 ABEJAS 040104  
 ABERRACION CROMOSOMICA  
 ABERRACIONES SEXUALES  
 CONDUCTA DE CAMPO ABIERTO ANIMAL 040402  
 ABOGADOS 140801  
 ABORTO ESPONTANEO 050105  
 ABORTO INDUCIDO 130307  
 ABORTO PROVOCADO  
 ABORTO TERAPEUTICO  
 ABSENTISMO 160502  
 ABSENTISMO ESCOLAR  
 ABSENTISMO LABORAL  
 ABSTINENCIA ALCOHOLICA  
 ABSTINENCIA DE DROGAS 090601  
 ABSTINENCIA DE TABACO  
 ABSTINENCIA SEXUAL 090501  
 DELIRIO POR ABSTINENCIA ALCOHOLICA 120112  
 SINDROME DE ABSTINENCIA 120208  
 ABSTRACCION 010404  
 PENSAMIENTO ABSTRACTO  
 RAZONAMIENTO ABSTRACTO  
 ABULIA 010402  
 ABUSO DE ALCOHOL 090602  
 ABUSO DE DROGAS 090602  
 ABUSO DE MEDICAMENTOS  
 ABUSO DE NIÑOS 120108  
 ABUSO SEXUAL 090802  
 APTITUD ACADEMICA  
 ESPECIALIZACION ACADEMICA  
 EVALUACION ACADEMICA 150303  
 FRACASO ACADEMICO  
 RENDIMIENTO ACADEMICO 150306  
 ACALCULIA 120209  
 ACCESO A LA INFORMACION 070103  
 ACCESOS REPETITIVOS DEL SUEÑO  
 ACCIDENTES CEREBROVASCULARES 120112  
 ACCIDENTES DE TRABAJO 160502  
 ACCIDENTES DE TRAFICO 091002  
 ACCIDENTES INDUSTRIALES  
 ACCIDENTES LABORALES  
 ACCION AUTOMATICA  
 ACELERACION DEL DESARROLLO  
 ACEPTACION DE LA TERAPIA  
 ACEPTACION DEL PACIENTE  
 ACEPTACION DEL ROL  
 ACEPTACION SOCIAL 090902  
 ACETILCOLINA 061001  
 ACIDO DESOXIRIBONUCLEICO 060101  
 ACIDO GAMMA AMINOBUTIRICO 061002  
 ACIDO RIBONUCLEICO 060101  
 DIETILAMIDA DEL ACIDO LISERGICO  
 ACIDOS 0601  
 ACIDOS NUCLEICOS 060101  
 ACNE 120510  
 ACOMODACION 080201  
 ACONTECIMIENTOS VITALES  
 ACOSO SEXUAL

## 2.6. Mantenimiento y actualización de tesauros

Dado el tiempo que requiere su elaboración, un lenguaje documental puede quedar anticuado incluso antes de que llegue a ser publicado, de forma que debe ser actualizado periódicamente.

El lenguaje del tesoro tiene, entre otras características, la de la flexibilidad, que permite ampliar regularmente su vocabulario.

Periódicamente, cada seis meses para un tesoro de nueva creación, y entre dos y tres años para aquellos que se han utilizado durante determinado período de tiempo, se procede a su revisión. Cuando ese lenguaje es especializado, la necesidad de revisión se acentúa, ya que cuanto más especializada es una materia, que requiere profundidad de indización, más rápidamente se vuelve obsoleta y se hace necesaria su puesta al día.

La actualización de un tesoro ha de hacerse tanto para incorporar la terminología derivada del desarrollo de la ciencia o materia a la que se dedica, como para cubrir lagunas y fallos detectados en la práctica de su aplicación a ese campo concreto del conocimiento.

Es necesario, pues, hacer un seguimiento del uso de los términos que componen el lenguaje documental para poder valorar cada una de sus entradas. Al indizar pueden aparecer conceptos en la documentación que no tienen cobertura en el vocabulario del tesoro. Cuando el indizador echa de menos un concepto, toma nota de la necesidad de un nuevo descriptor, que registra o incluye en una lista de espera o en un fichero de términos candidatos a descriptor (véase figura 13) que revisará y analizará en el momento de la actualización. Si varias instituciones comparten el uso del tesoro, cada una

aportará fichas o listas de términos candidatos, que en el momento de la puesta al día deberán ser analizadas con detenimiento para que los términos incluidos, excluidos o simplemente modificados, resulten operativos para la mayoría de los centros implicados, evitando en la medida de lo posible la permanencia en el tesoro de conceptos o incluso campos de uso exclusivo en una sola institución.

La inclusión de los nuevos descriptores no puede realizarse día a día porque crearía confusión y rompería la estructura característica del tesoro. Hay que tener en

**ANVERSO**

<b>DESCRIPTOR:</b>		<b>FACETA O CAMPO:</b>
<b>NOTA ACLARATORIA:</b>		
<b>USADO POR:</b>		
<b>TÉRMINO GENÉRICO</b>	<b>TÉRMINO ESPECÍFICO</b>	<b>TÉR. RELACIONADO</b>

**REVERSO**

<b>OBSERVACIONES:</b>
<b>TERMINO PROVISIONAL:</b>
<b>FECHA DE INCLUSIÓN:</b>
<b>FECHA DE REVISIÓN:</b>
<b>FECHA DEL DESCARTE DEL TERMINO:</b>

Modelo para el registro de descriptores

Figura 13

cuenta que, cuando cambiamos un término, debemos cambiar a su vez todas las relaciones de ese término dentro del tesoro.

La revisión es, por lo tanto, un proceso continuo, que permite conocer el uso que se hace de la terminología tanto en la operación de indización como en la de recuperación de información, porque al igual que en las operaciones de entrada, las consultas realizadas en lenguaje natural por los usuarios aportan la terminología real del centro de documentación o sistema documental. Las ecuaciones de búsqueda nos dan el porcentaje de precisión y respuesta que conseguimos con los descriptores.

La presencia de sinónimos y cuasisinónimos en las consultas debe tenerse en cuenta, haciéndose necesaria su incorporación porque favorece el acceso del usuario a la información bajo diferentes denominaciones de un mismo concepto.

Van Slype<sup>15</sup> reconoce la importancia de conocer la distribución estadística de los *descriptores para fundamentar el mantenimiento del tesoro*. Considera que, según muestra la experiencia, la frecuencia de utilización de los términos es muy diversa: gran cantidad de descriptores no se usan o se usan muy poco; una pequeña cantidad de descriptores se utilizan con gran frecuencia, en general el 20 por ciento de los descriptores más usados representan el 80 por ciento de las indizaciones. Chaumier<sup>16</sup> también hace referencia a la discordancia existente entre el empleo de términos en el momento de la entrada de documentos al sistema como su empleo en las ecuaciones de búsqueda; es por ello importante analizar la terminología utilizada por el usuario, que se reduce por lo general a un restringido vocabulario.

Al igual que en el proceso de elaboración, en la actualización se tendrán en cuenta:

- a) El grado de evolución del lenguaje
- b) La opinión de los especialistas en la materia
- c) Las consultas de los usuarios
- d) La utilización de los descriptores en el proceso de indización.

---

<sup>15</sup>SLYPE, G.van. Op. cit., p. 109.

<sup>16</sup>CHAUMIER, J. Op. cit., p. 128.

Pueden darse cambios de las siguientes características:

1. Inclusión de términos nuevos. En el momento de la revisión se comprueba la frecuencia de uso de estos términos, que previamente se registraron en la lista de espera mencionada anteriormente, y la posibilidad de su inclusión en el tesauro como descriptor definitivo.

2. Eliminación de un descriptor. Aquellos términos incorrectos, poco o nada usados, se pasan a la categoría de no descriptores, pero no se eliminan, porque pueden ser fuentes de acceso al tesauro para los usuarios. Se les añade la sigla *USE* y se mantienen en el índice, previa anulación de todas sus relaciones. Ej.:

Zonas costeras

**USE ZONAS LITORALES**

3. Subdivisión de un descriptor presente en el tesauro en una serie de descriptores de menor nivel jerárquico. Es decir, se sustituye un descriptor que figura en el tesauro como término genérico por varios términos más específicos, incluyendo un nivel de especificidad mayor en el tesauro. Por ejemplo, reemplazamos MINERALES por CROMO, ESTAÑO, HIERRO, LITIO, porque se ha detectado la necesidad de utilizar conceptos más específicos al aplicar el tesauro a un fondo documental determinado. En este caso puede eliminarse el término más general del que derivan los otros, o puede mantenerse como término genérico incluyendo a todos los demás. Ej.:

TG MINERALES

TE CROMO

ESTAÑO

HIERRO

LITIO

Esta opción resulta adecuada cuando se trata de descriptores de uso muy frecuente. Como alternativa a la subdivisión del término podría añadirse una palabra que especifique más su sentido. Ej.:

DISTRIBUCIÓN DE ENERGÍA

DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA

DISTRIBUCIÓN DE ENERGÍA HIDRÁULICA

4. Inclusión de descriptores de un grupo jerárquico en otro. Por ejemplo, el descriptor MANO DE OBRA podría figurar en un tesoro en el campo semántico RECURSOS HUMANOS y pasar, tras la actualización, a formar parte del grupo jerárquico POBLACIÓN ACTIVA, si se considera conveniente.

5. Cambio de términos sinónimos. Un término considerado no preferente o no descriptor en el tesoro, pasa a ser preferente. Puede haber descriptores cuyo significado es demasiado próximo, creando duda al indizador en el momento de su elección para describir un concepto, o que deben ser incluidos simultáneamente en el momento de la formulación de las consultas. En ambos casos se analizará la conveniencia de mantenerlos o de modificarlos. Ej.:

MANDIOCA

UP Yuca

6. Si se detecta la existencia de términos homónimos, se procede a su desambiguación. Ej.:

RAÍZ

TR MATEMÁTICAS

7. Incorporación o eliminación de relaciones semánticas. La inclusión o exclusión de alguna relación implica igualmente la modificación de su correspondiente referencia cruzada: *TG-TE*, *TR-TR*.

En la nueva edición revisada deben aparecer los cambios realizados en el tesoro en relación con la edición anterior. Para facilitar el acceso a la información, independientemente de los cambios de denominación sufridos por algunos términos, conviene especificar mediante notas históricas la fecha en la que determinado descriptor ha sufrido modificaciones.

Con la actualización surge el problema de la correspondencia entre los términos incluidos, excluidos o modificados en la nueva edición, y la notación de los documentos existentes con anterioridad a la puesta al día del tesoro. Existen posibles soluciones para resolver la falta de correspondencia entre las entradas del tesoro y los documentos indizados:

1. Acordar la utilización de los nuevos descriptores únicamente para recuperar los documentos tratados después de la introducción de los mismos en el tesoro, lo que figurará en nota histórica (NA) junto a cada uno de los términos afectados. En la misma nota se informará al usuario de los descriptores que existían con anterioridad bajo ese tópicico, con objeto de posibilitar su búsqueda. Este procedimiento facilita la labor de indización pero dificulta la búsqueda que, a pesar de las notas aclaratorias, resulta más complicada. Ej.:

MINERALES (CROMO 10/10/97; antes úsese MINERALES)

2. Volver a indizar los documentos almacenados. Esta posibilidad se puede poner en práctica paulatinamente: a medida que se va indizando con los nuevos descriptores, se recuperan y actualizan los documentos existentes bajo su anterior denominación. Aunque esta opción es la menos popular, por el esfuerzo que implica, es quizá la más correcta. Van Slype<sup>17</sup> cree necesario evitar, en la medida de lo posible, suprimir descriptores que hayan sido ya utilizados para indizar los documentos; si esta supresión es absolutamente necesaria (como en el caso de un descriptor demasiado frecuente) será necesario reindizar todos los documentos en los que esos descriptores intervienen, para mantener la coherencia entre el fichero de búsqueda y el tesoro.

Como conclusión de este capítulo podríamos decir que el tesoro es un lenguaje documental que, por las características señaladas, parece idóneo para la recuperación de información en bases de datos bibliográficas. En el siguiente apartado vamos a analizar los primeros pasos de su desarrollo y empleo en España, teniendo siempre presentes las

---

<sup>17</sup>SLYPE, G. van. Op. cit., p. 109.

bases conceptuales señaladas, dado que ha existido y existe la tendencia a aplicar la denominación de tesoro a otros tipos de lenguajes que no responden al perfil aquí descrito.

## Capítulo 3

# Orígenes y evolución de los tesauros en España

### 3.1. Antecedentes históricos del tesoro

La organización de vocabularios fundada en bases conceptuales es una idea que, de una u otra forma, ha estado presente en la mente de todos aquellos que se interesaron en organizar las distintas ramas del saber a lo largo de la historia. Esta idea ha aparecido en diferentes contextos, para diferentes propósitos, y los argumentos empleados con este fin, coinciden a grandes rasgos con los planteados en relación con la naturaleza y construcción de tesauros documentales.

La literatura existente acerca del origen de los tesauros no coincide en la apreciación de determinados factores como determinantes del uso y aparición de los mismos con fines documentales. Nosotros nos hacemos aquí eco de algunas de las principales tendencias en este sentido.

El origen primero del término tesoro proviene de la palabra griega *θησαυρος*, con el significado de colección. El latín adoptó el vocablo con el sentido de tesoro y, posteriormente lo utilizaron las lenguas modernas para referirse, en términos generales, a repertorio o recopilación de palabras o colecciones de una o varias áreas del conocimiento.

La primera persona que utilizó la expresión *thesaurus* fue el florentino Bruneto Latini (1220-1294) en su enciclopedia *Livre dou Tresór*, escrita en francés entre 1262 y

1268. En el siglo XV se tiene constancia de su uso por Cirilo de Alejandría en su *Thesaurus adversus haereticus*<sup>1</sup>.

Tanto en la Edad Media como durante el Renacimiento la denominación de *thesaurus* se restringía a los diccionarios monolingües, como, por ejemplo, el *Thesaurus Linguae Latinae* (1532) (figuras 14 y 15), y el *Thesaurus graecae linguae*, de Henri Estienne (1572).

En España, la primera vez que aparece escrita la palabra *thesaurus* fue en la obra *Compilatio Maior o In excelsis Dei Thesaurus*, compilación del derecho aragonés que realizara el jurista catalán Vidal de Cañellas (m. en 1252).

El *Diccionario de autoridades* de la Real Academia Española lo define como: *Índice, ò compendio de voces, y términos reducidos de una Lengua à otra*<sup>2</sup>. El *Diccionario de Covarrubias*<sup>3</sup> lo interpreta como tesoro en su más estricto sentido material, es decir, como conjunto escondido de monedas o bienes materiales.

---

<sup>1</sup>LAFUENTE CASTAÑEIRA, I. *El concepto de tesoro en España ayer y hoy: 1970-94*. Proyecto de investigación dirigido por la Dra. Emilia Currás. Curso de Especialista Universitario en Documentación Pedagógica 1994/95. Facultad de Ciencias de la Educación. Universidad Complutense de Madrid.

<sup>2</sup>DICCIONARIO de la lengua castellana, en que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las frases o modos de hablar, los proverbios o refranes, y otras cosas convenientes al uso de la lengua. Madrid: En la Imprenta de Francisco del Hierro, Impresor de la Real Academia Española, 1739.

<sup>3</sup>COVARRUBIAS OROZCO, Sebastián de. *Tesoro de la lengua castellana o española*. Madrid: Por Luis Sánchez, 1611: TESORO. *Latine* THESAURUS, *a graeco* θησαυρός; un escondidijo y lugar oculto, do se encerró alguna cantidad de dinero, oro o plata, perlas y joyas y cosas semejantes de tanto tiempo atrás que dello no había memoria ni rastro alguno, ni de quién fuese; de donde se sigue tener justo derecho a ello el que se lo halla, como cosa que no es de otro ni puede decir ninguno que es suyo. Cerca desta materia disponen las del reino lo que se deba hacer, estando obligado el que halla el tesoro a manifestarlo, y dar al rey la parte le cabe. 2. También llamamos tesoro el dinero o joyas de un príncipe o de una república que está en custodia, *a verbo --código, recondo, repono*. Es hebreo, que *tesurah*, תְּסוּרָה, vale tesoro. Algunas veces significa tesoro el lugar abditísimo [sic: abditissimus] y profundo, como se entiende de aquel verso del Salmo 134: "*Qui produit ventos de thesauris suis, id est, de abditissimis et absconsis cavernis terrae*". Tesoro de duende, suelen decir que los duendes tienen escondidos tesoros, y, cuando alguno los halla, volvésele en carbones, de do nació el proverbio: *Thesaurus carbones facti sunt*.

# THESAURVS

L I N G V A E

L A T I N A E

S E V

P R O M P T V A R I V M D I C T I O N V M E T  
*loquendi formularum omnium ad Latini sermonis perfectam notitiam assequendam pertinentium: ex optimis auctoribus concinnatum.*



C V M P R I V I L E G I O C A E S A R E A E M A -  
I E S T A T I S E T C H R I S T I A N I S S . G A L -  
L I A R V M R E G I S .

L V G D V N I .

*Ex libris Moredani  
Lxi.*



LITTEra vocalis, prima in Alphabeti serie, quæ sono infuauiffima est, si sapius uteretur in fine. Cicero in Orat. Aurtus artes Colchorum splendidi nominibus illuminatus est versus: sed proximus inquinatus infuauiffima litera finitus, Irugifera & fertula arua Afia tenet. Aflicat reliqua literæ fua figura pingituz. Cic. de Diuin. Sus roftro fi homi A lectram imprefferit, num propterea fufpicari poteris Andromachâ Ennij ab ea polle defcribè

A, litera nota erat in fuffragiis & comitus legum qua fe ciues repudiare intentionem fignificabant. Nam duj dari tabulæ foletant, quarum in altera V. R. ad efl. Vi. rogat, fcriptum erat in altera A. ad efl. Antiquo, antiqua fequor, noua non placet.

Deiur item A, litera falutaria à Cicerone in oratione pro Milquod vt ferbit Afconius, fi videbatur iudicibus bona caufa, abfoluunt reu, notamque abfolutionis A, literam in tabella, quam à prætorè accipiebant, fcribebant: quemadmodum contra, vltimè literam eodem loco appellat literam C, quod vt ibidem fcribit Afconius, condemnadi nota efler. Cic. pro Mil. Quod nifi vidiffet, polle abfolui tum qui faeretur: cum videret non faerit, neque queri inquam iuffiffet, nec vobis tam falutare hanc in iudicando literam, quam illam triffem dediffet.

A, puncto poftpofto, prænomen fignificat, vt A. Cecinna id efl, Aulus. A (ficut reliquæ literæ) ponitur interdum fola pro fola dictione cuius efl initium, vt A. A. And efl. Arg. Argentio, Auro: quo de genere feri pei funt à grammaticis integri libri. A, præpoftionem Latinam deduci putant à Græca præpoftione α, detracta vltima fylaba.

Abentiam inde ductam tradunt. Si quidem à litera vltima eliditur ab x, præferunt in colliſionibus vocalium: id efl fequente dictione à vocali incipere, & remanet α. Quod fi e tenuem veteris in A fuam incipiam, habebis Ab præpoftionem Latinam.

A & Ab, præpoftiones videtur factæ ex Ab præpoftione. Cic. in Orat. Vna præpoftio efl ab, eaque nunc tantum in accepti tabulis manet, ne his quidem omnium in reliquo fermone mutata efl. Nam Amouit dicimus, & Abegit, & Abſtulit: vt iam nefcimus, Ab ne veram fit, an Abi. Quid fi eum Abfugi turpe vifum efl, & Abfer noluerunt, Aufer maluerunt. Quæ præpoftio, præter hæc duo verba, nullo alio in verbo reperitur. Hactenus Cæſar: vbi demonſtrat Abi præne exoleu, & Abi in cõpoſitione duorum tarum verboru, Abfuppo, & Abferre inueniunt. Significat autem præpoftiones Abi, Ab, A, caufam efficiendam: id efl feruunt verbis paſſiue fignificationis in ablatiuo.

Abi, videtur eſſe ex Græco α, retro. Vnde & Feſto retractationem fignificat. Ter. Adel. -abi quiuus homine. Idem Phor. Vt abi te abij. Quibus autem literis præponatur, compoſitio indicat, vt Abſcondo, Abſtineo, &c.

Ab & A frequentioris funt vfus. Ab corripitur. A producitur. Ab præponitur omnibus literis, A folis conſoni. Quintil. lib. 12. cap. 1. Fabritius reſpondit à ciue fe ſpoliari male, quam ab hoſte tenere. Horat. Serm. ſayr. 1. Laudatur ab his, culpatur ab illis. Verri poſſit hic ablatiuus in dantiu, Græco more, vt ſeruo placet. Virgil. Eneid. 1. Neque cernitur. vll. Id efl. ab vilo. A meis ciuibus ea res mihi fua honoris, id efl, Ob eam rem honoribus affectus ſum à ciuibus meis. Cic. pro Milone, Beatos eſſe quibus ea res honoris fuerit à ſuis ciuibus. Idem pro Rabirio Poſth. Quorum impunitas fui non modo à iudicio, ſed etiam à fermone. Idem lib. 7. ad Att. Sumus enim flagitioſe imparati cum à militibus, tum à pecunia. Idem lib. 3. de Orat. Hæc cum Craſſus dixiſſet, parumper & ipe conuenit, & à cæteris ſilentium fuit. Abſtamen codices præpoftionem non habent. Plaut. Epid. 1. A morbo valui, ab animo æger fui. Idem Curc. 1. Tam à me pudica eſt, quaſi furor mea fit. Terent. Heccyra, 1. Narratque vt virgo à ſe integra etiam tum ſet. Cic. Att. Cum Romæ à iudiciis forum reſcrixit.

A Platone & Ariſtotele homines, id eſt, Academiici & Peripatetici, ſiue Platonicæ & Ariſtotelici, Platonis & Ariſtotelis doctrinæ ſequentes. Cic. pro Mur. Noſtri inquam illi à Platone & Ariſtotele moderati homines & temperati, aiunt apud ſapientem valere aliquando gramam.

Ab aliquo legare. Cic. pro Cluentio, Itaque ei teſtamento legat grandæ pecuniam à filio, ſiqui natus eſſet: ab ſecundo hæredè nihil legat. Idè ad Atticam libr. 13. ep. 45. Cognoui Clunium à T. Hordeonio legare & Terentiu n. 1. 2. 2. & ſepulcro, nulliſque rebus, nihil à nobis. Id eſt, Cognoui Clunium iubere, vt Hordeonius cohæres meus Terentiu numeraret, & ad ſepulcrum ædiſcandam, nulliſque aliis reſ ſumptus ſuppeditaret: ego verò nihil darem, ita vt meliore conditione ſim quam Hordeonius cohæres meus. Illi enim de ſua parte detrahendum eſt, quod Terentiu ſoluat, quod ſepulcro nulliſque rebus, mihi proſus nihil.

Ab aliquo poſſidere. Vlp. in l. j. §. 5. D. Vt poſſideris. Quod ait Prætor in interdico, Nec vi, nec clam, nec præcaro alter ab altero poſſideris, hoc eò pertinet, ſi qui poſſidet vi, aut clam, aut præcaro: ſi quidem ab alio, proſit ei poſſeſſiu: verò ab aduerſario ſuo, non debeat cum propter hoc, quod ab eo poſſidet, vincere. Paulus in l. iij. in princ. eo. ut. Ego poſſideo ea iuſta cauſa: vi, aut clam. Si à me poſſides, ſuperior ſum in interdico: ſi verò non à me, neuter noſtrum vincetur, nam & tu poſſides, & ego. Ab alio poſſidere vi, aut clam, aut præcaro, eſt poſſeſſionem vſurpate rei alienæ aut vi, aut clam, aut præcaro. Id eſt, cum alius poſſideret, tu illum poſſeſſione deieciſti aut vi, aut clam, aut præcaro.

Ab aliquo via vi. Vlpian. in l. iij. in princ. D. De itinere ad ique priuato. Si cum à me redè via vrebatur, fundum vendidero per quem vrebatur: deinde emptor te prohibuerit, licet clam videaris ab eo vi (nã qui prohibitus vinitur, clam vinitur): iamen interdico tibi competere in ea anuum. Paulo poſt: Si quis ab auctore meo vi, aut clam, aut præcaro vſus eſt, redè à me via vi prohiberet, & interdico tibi in idem eſt: quia à me viderit vi, aut clam, vel præcaro poſſidere, qui ab auctore meo vicioſe poſſidet. Nam & Pedius ſcribit: Si ſtans clam, aut præcaro ab eo vſus ſe, non in cuius locum hæreditate, vel emptione, aliòve quo iure ſucceſſiſſem eſſe dicendum. A me via vi eſt, cum ego Dominus eſſem fundi, tibi conceſſi, vt via in meo fundo vtereris. Ab auctore meo vi, cum auctori meus qui mihi fundum vendidit, tibi eam viam conceſſit.

Ab altero ſeu. Vlpian. in l. iij. in princ. D. De ſuperficiebus. Alter ab altero fruamini. Id eſt, altero alteri fruamini concedente.

Ab aliquo aquam ducere, id eſt, per alicuius fundu in ſuum deductam habere. Vlpian. in l. j. §. Deinde ait prætor. D. de aqua quotidiana & æſtina. Vt priore ætate aquam, qua de agru, nec vi, nec clam, nec præcaro ab illo ductiſti, quo minus ita ducas vni fieri videri, id eſt, vt per illud fundum in tuum ductiſti.

Conſimilia etiam ſunt ſequentia. Ab aliquo ſoluere, dare, numerare. Cic. Att. lib. 1. Non modò non recuſare, ſed etiam hoc dicere, ſe à me ſoluere. Quod enim Prætori dare conſueſcerent, quoniam ego non acceptam, ſe à me quodammodo dare. Hoc eſt, quaſi de meo, vel de mea pecunia. Cic. Att. l. 7. ep. 16. Vtæ laborat, vt tibi quod debet, ab Egnatio ſoluat. Id eſt, ſumptam ab Egnatio debitor ſuo pecunia tibi numeret. Idem pro Flacco. Si Prætor dedit, vt eſt ſcriptum, à Quæſtore numerat: Quæſtor à menſis publicæ: menſis aut ea veſtigia, aut ex tributis. A Quæſtore, inquit, quaſi à Quæſtore acceptam, intellige pecuniam.

Perſoluere ab aliquo. Cic. Sinite me, quod vobis ſeruas, hæc ſæpe promiſſi, id à vobis ſi perſoluere.

Repreſentare pecuniam ab aliquo. Cic. Att. 12. Reliquæ pecuniam vel vſuram ſilio pendemus, dum à Faberio, vel ab aliquo, qui Faberio debet, repreſentabimus. Quaſi dicat, Spero fore, vt à Faberio exigamus ita cito, vt poſſimus ſilio ſoluere ante diem conſtitutam: ſic quo pendemus vſuram, dum repreſentemus.

Suppeditare ab aliquo. Cic. contra Rullum, Sumptum horis ex ætario ſuppeditat à ſociis.

A cauda de ovo cauro. Plin. lib. 10. cap. 16. Hylas tradit noctuam, bubonem, picum arboris cauſantem, trogonem, cornicem, à cauda de ovo cauro: quoniam pondere capitiſſi percuſa ova, poſteriorum partem corporum fouendam manui applicant. In ſimilibus tamen locutionibus interdum ſupprimunt præpoſitio: vt apud eundem lib. 7. cap. 8. Neroneum pedibus genium patens eius ſcribit Agrippina. Ratu nauem capite hominem gigni moſ eſt, pedibus efferri. Item mutatur in præpoſitionem in aliquando. Plin. lib. 10. cap. 24. Quætanque ait, animal parium, in capis gignunt, circum acto ſub enixam ſeruo, aliis in vtero poſtecto.

A matre pulli: id eſt, recentes à matre accepti. Colum. lib. 8. cap. 5. A matre pulli non exiguis pecuis vntem.

Elephantu nuper à ſylua. Flor. lib. 4. cap. 1. Elephanti bellorum rudes & nuper à ſylua, conſternati ſubito clangore, &c.

A lenone manu afferere. Plaut. Perſa, 4. Continuo in illam à lenone aſſerit maan. Sermo eſt conſiſus, pro eo quod eſt, à lenoni poſſeſſione vel dominio.

Ab exale propè eſſe. Tacit. lib. 17. Licinius Mutianus, vit ſecudis aduerſique luxu famoſus, ſuſpecta Claudiu iracundia in ſecretum Afæ reſponſus, iam propè ab exale ſui, quam poſtes à principe.

A contumelia quàm à laude propius fuerit, poſt Vitellium eligi. Tacit. lib. 18. Hoc eſt, Propius contumeliæ quàm laudi, Contumeliouſam magis quàm glorioſam.

Ab amore ſcribere. Balbus Cic. lib. 9. Hæc quàm prudenter tibi ſcribam, nefcio: ſed illud certe ſcio, me ab ſingulari amore beneuolentiſſimo quecuoque ſcribo tibi, ſcribere. Gallicè dicimus, De grand amore.

Ab aſpectu decoro: pro eo quod Galli dicimus, Beau à voir. Colum. lib. 6. cap. 1. Apenninus boues progenerat duriffimos, omnemque difficultatem tolerares, nec ab aſpectu decoroſe. Cicero dicit, Ad aſpectu præclaro, Ad aſpectum pulchro. Vide ſuis locis.

Ab aliquo deſcribere. Cic. lib. 13. ad Attic. Scripſit Balbus ad me, ſe à te quoniam de Finibus librum deſcripſiſſet, id eſt, ex tuo libro.

A ſe aliquid promere. Cic. de Orator. lib. 1. Videat non ipſe aliquid à me promiſſe, ſed fortuò in ſermonem veſtrum incidiſſe.

A ſe aliquid facere, Cic. de Orat. 1.

A me ſalutem die Ciceroniid eſt, verbis meis, ſeu nomine meo. Cic. ad Attic. lib. 6. Valebis igitur, & valere Filiam & Ceciliam noſtris iubebis literis, & ſalutis à meo Cicerone.

A me illos habe commendatiſſimos, Cicero lib. 12. ep. 10. 1. ad Cornificium.

A ſole calor. Cic. de Nat. deor. lib. 2. Gallina pullos cum exenderint, ita tumentur, vt & pennis foucant, ne à frigore lædantur: & ſi eſt calor à ſole, ſe opponant.

A ſirpe, id eſt à maioribus. Salluſt. in Jugurth. Jugurtha homo omnium quos terra ſullinet ſceleratiſſimus, contempto imperio veſtro, Maſſinæ me nepotem, etiam à ſirpe ſocium atque amicum populi Romani, ægno fortunique omnibus expulſi: Gallicè, De ſirpe.

A ſirpe, pro ſirpius & iudicis. Lia. 4. bel. Maced. Equidem fabulam à ſitam rem ducebam eſſe, viroorum omne genus in aliqua inſula conuinatione muliebri à ſirpe ſublucam eſſe. Ab auro fulgoreto reuereri. Lucret. lib. 2. 1. 1. Auda èt que inter reges, ætümque potentis Verſantur: neque fulgorem reſerentur ab auro.



El *Shorter Oxford English Dictionary* ofrece en 1736 la siguiente definición del término: *un tesoro o depósito de conocimiento, como un diccionario, una enciclopedia o similar*. El diccionario de la Real Academia Española lo delimita en términos parecidos: *tesoro, nombre dado a algunos diccionarios, catálogos, antologías, etc.*

Aunque el uso de la palabra *thesaurus* tuvo diversas interpretaciones, -casi todas coincidentes con su consideración como repertorio, diccionario, compendio de términos o vocabulario- durante los siglos XVII y XVIII muchas obras litúrgicas llevaban esta denominación. El diccionario enciclopédico Espasa<sup>4</sup>, se hace eco de dicha utilización con la siguiente definición: *Muchas obras de erudición profana y eclesiástica llevan este título, unas son diccionarios, otras colecciones de disertaciones o escritos de autores diversos. Se relacionan con la liturgia:*

1º. Bartolomé Gavanti, *Thesaurus Sacrorum Rituum*, cuya primera edición fué publicada en 1628.

2º. Suicer, *Thesaurus ecclesiasticus*

3º. Martène y Durand, *Thesaurus novus anecdotorum* (Paris, 1717). Se trata de una especie de complemento a la Patrología. Contiene importantes textos de escritores de la Edad Media relativos especialmente al monacato.

4º. Antonio Zaccaria, *Thesaurus theologiarum disertationum* (Venecia 1762).

---

<sup>4</sup>ENCICLOPEDIA universal ilustrada europeo americana. Madrid: Espasa Calpe, 1928.

5°. A. Daniel, *Thesaurus hymnologicus sive hymnorum, canticorum, sequentiarum circa annum MD usitatorum collectio amplissima. Carmina collegit, apparatu critico ornavit, veterum interpretum notas selectas, suasque adiecit.* (Halis, 1841-56).

6°. Voldebing, *Thesaurus commentationum* (Leipzig, 1884).

Pero el uso de la palabra thesaurus para referirse a diccionarios de sinónimos parece ser, en definitiva, el que establece mayores conexiones con su futuro uso en sistemas documentales, por lo que, para analizar la evolución histórica de este lenguaje documental, será conveniente proceder al examen de tales diccionarios. De hecho el control de la sinonimia, así como la agrupación de palabras relacionadas, ha estado siempre presente en la mente de los compiladores de vocabularios generales o especializados y, por supuesto, es una de las características a destacar de las primeras clasificaciones bibliográficas y de las listas de encabezamientos de materias elaboradas a finales del siglo pasado.

Jones<sup>5</sup> considera relevante el tratamiento de la sinonimia lingüística en la historia de los tesauros, así como en vocabularios organizados en base a la agrupación de palabras relacionadas. *La más antigua y persistente forma de clasificación del vocabulario -dice esta autora- es el thesaurus cuya finalidad es contribuir a la descripción del lenguaje y a servir de ayuda en su uso. Añade que la clasificación por materias de vocabularios, la agrupación de palabras relacionadas, se remonta al Sanscrit Amara Cosha, al clásico Onomastikon de Pollux, y al Aelfric's Anglo-Saxon Latin Glossary. En el Onomastikon, las palabras, incluidos los sinónimos, se agrupan bajo diez encabezamientos como Dios, el Hombre (partes del cuerpo), Ciencia y Arte, Caza y Comidas. En*

---

<sup>5</sup>JONES, K.S. Some thesauric history. *Aslib Proceedings*, v. 24, n° 7 (1972) p. 400.

*Aelfric* las palabras están organizadas bajo encabezamientos tales como *Herramientas para la Agricultura, Ocupaciones, Insectos, Ropas y Bebidas*. Existe, en esta forma de presentación, un notable paralelismo con los tesauros documentales, que agrupan semánticamente la terminología que los compone.

En 1816 Crabb editó su *English Synonymes Explained in Alphabetical Order with copious ilustrations and examples* (figura 16), que fue el primero en expresar la información no sólo en forma alfabética sino adoptando un orden conceptual. Esta forma de presentación le convierte en precursor del *Roget's Thesaurus of English words and phrases*, que, tradicionalmente, se considera precursor de los tesauros destinados al tratamiento de la información en sistemas documentales.

Peter Mark Roget nació en 1779, y en 1848, ya retirado de su vida profesional, se dedicó por entero a la tarea de preparar, para su publicación, la obra que perpetuaría su memoria. Se trataba de un catálogo de palabras organizadas según sus significados, en cuya compilación había estado trabajando desde 1805. La primera edición apareció en 1852 con el título *Thesaurus of English words and Phrases Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary Composition*.

El *Thesaurus de Roget* no es, por tanto, un simple diccionario de sinónimos, sino una clasificación ideológica del vocabulario. Se caracteriza, fundamentalmente, por la forma de presentación de sus entradas. Hasta su aparición, los diccionarios organizaban las palabras en orden alfabético, a excepción, como hemos mencionado, del *English Synonymes Explained in Alphabetical Order*, de Crabb. El *Roget's Thesaurus* crea una cadena de sinónimos que forman categorías conceptuales, presentando las palabras de acuerdo con las ideas que expresan .... *la palabras y frases del lenguaje están aquí*

idea common to these terms, but the former is employed more properly against the person; the latter against the thing.

*Abuse* is addressed to the individual, and mostly by word of mouth; *invective* is communicated mostly by writing.

*Abuse* is dictated by anger, which throws off all constraint, and violates all decency; *invective* is dictated by party spirit, or an imperate warmth of feeling in matters of opinion.

*Abuse* is always resorted to by the vulgar in their private quarrels; *invective* is the ebullition of zeal and ill nature in public concerns.

The more rude and ignorant the man, the more liable he is to indulge in *abuse*; the more restless and opinionated the partisan, whether in religion or politics, the more ready he is to deal in *invective*.

We must expect to meet with *abuse* from the vulgar whom we offend, and if in high stations our conduct will draw forth *invective* from busy bodies whom spleen has converted into oppositionists.

At an entertainment given by Plautus to some of his intimates, Thralippus, a man of violent passion and infamed with wine, took some occasion, not recorded, to break out into the most violent *abuse* and insult. COMMERSLAND.

This is the true way of examining a libel; and when men consider that no man living thinks the better of their heroes and patrons for the panegyric given them, none can think themselves lessened by their *invectives*. STEELE.

#### TO ACCEDE, CONSENT, COMPLY, ACQUIESCE, AGREE.

ACCEDE, in Latin *accedo*, compounded of *ac* or *ad* and *cedo* to go or come, signifies to come or fall into a thing.

CONSENT, in French *consentir*, Latin *consentio*, compounded of *con* together and *sentio* to feel, signifies to feel in unison with another.

COMPLY comes probably from the French *complaire*, Latin *complaceo*, signifying to be pleased in unison with another.

ACQUIESCE, in French *acquiescer*, Latin *acquiesco*, compounded of *ac* or *ad* and *quiesco*, signifies to be easy about or contented with a thing.

AGREE, in French *agr  er*, is most probably derived from the Latin *gruo*, in the word *congruo*, signifying to accord or suit.

We *accede* to what others propose to us by falling in with their ideas; we *consent*\* to what others wish by authorising it: we *comply* with what is asked of us by allowing it, or not hindering it; we *acquiesce* in what is insisted by accepting it and conforming to it; we *agree* to what is proposed, by admitting and embracing it.

We object to those things to which we do not *accede*; we refuse those things to which we do not *consent*, or with which we will not *comply*; we oppose those things in which we will not *acquiesce*; we dispute that to which we will not *agree*.

To *accede* is the unconstrained action of an equal; it is a matter of discretion: *consent* and *comply* suppose a degree of superiority, at least the power of preventing; they are acts of good-nature or civility: *acquiesce* implies a degree of submission, it is a matter of prudence or necessity: *agree* indicates an aversion to disputes, it respects the harmony of social intercourse.

Members of any community ought to be willing to *accede* to what is the general will of their associates. Parents should never be induced to *consent* to any thing which may prove injurious to their children. People ought not to *comply* indiscriminately with what is requested of them. In all matters of difference it is a happy circumstance when the parties will *acquiesce* in the judgment of an umpire, which is the greatest proof of their willingness to *agree*.

At last persuasion, menaces, and the impending pressure of necessity, conquered her virtue, and she *acceded* to the fraud. COMMERSLAND.

My poverty, but not my will *consents*. SHAKESPEARE.

Inclination will at length come over to reason, though we can never force reason to comply with inclination. ARNDSON.

This we ought to *acquiesce* in, that the Sovereign Being, the great Author of Nature, has in him all possible perfection. ADDISON.

We *agreed* to adopt the infant as the orphan son of a distant relation of our own nation. COMMERSLAND.

#### ACCENT, EMPHASIS, STRAIN, STRESS.

\* "ALL these words denote an increased effort of voice.

"ACCENT (which is derived from *cantus* a song) describes that sort of exertion which varies the utterance from low to high, from grave to acute, from flat to sharp, from hoarse to shrill. EMPHASIS (which is derived from *vis*: to indicate) describes that sort of exertion, which varies the utterance from soft to loud, from quick to slow, from faint to marked, from slurring to distinct. STRAIN is the English word for accent. STRESS is the English word for emphasis. *Strain* is derived from the Saxon *streng* strength, and means a strengthening of the voice; *stress* is the substantive of the verb to *stretch*, which is the intensive of *to stride*; it means therefore a stride of voice which excludes the idea of gradual or musical intonation."

\* As the rise and fall of sound prevents monotony, which would give a deadness to the human speech, accent is not improperly called in Diomedes *anima vocis*.

FOSTER on Accent and Quantity.

\* It may be remarked that *accent*, though closely united with quantity, is not only distinct from it, but in the formation of the voice really antecedent to it. The pitch or height of the note is taken first, and then the continuance of it is settled: by the former of these *accent* is determined, by the latter quantity."

FOSTER on Accent and Quantity.

\* *Accent*, in the Greek names and usage, seems to have regarded the tone of the voice; the acute *accentus* raising the voice in some certain syllables to a higher or more acute pitch or tone, and the grave depressing it lower. HOLDEA.

\* *Emphasis* not so much regards the time as a certain grandeur, whereby some letter, syllable, word, or sentence, is rendered more remarkable than the rest by a more vigorous pronunciation, and a longer stay upon it. HOLDEA.

Though long and short, or short and long syllables, may sometimes form the rhythm of English verse, yet that which invariably and essentially forms it is the interchange of *emphatic* and *non-emphatic* syllables. BEATTIE.

Our words flow from us in a smooth continued stream, without those *strainings* of the voice, notions of body, and majesty, which are so much celebrated in the orators of Greece and Rome.

ATTRENAVY.

Throughout 'the Gothic Dialects the stress falls on the radical syllable. TAYLOR.

as to leave no power of giving a particular force to other words, which though not equally, are in a certain degree *emphatical*, and placing the greatest *stress* on conjunctive particles and other words of secondary importance. KURIMIA.

#### ACCEPTABLE, GRATEFUL, WELCOME.

ACCEPTABLE signifies worthy to be accepted.

GRATEFUL, from the Latin *gratus* pleasing, signifies altogether pleasing; it is that which recommends itself; the *acceptable* is a relative good; the *grateful* is positive; the former depends upon our external condition, the latter on our feelings and taste: a gift is *acceptable* to a poor man, which would be refused by one less needy than himself; harmonious sounds are always *grateful* to a musical ear.

WELCOME signifies come well or in season for us.

*Acceptable* and *welcome* both apply to external circumstances, and are therefore relatively employed; but the former is confined to such things as are offered for our choice, the latter refers to whatever happens according to our wishes; we may not always accept that which is *acceptable*, but we shall never reject that which is *welcome*. It is an insult to offer any thing by way of a gift to another which is not *acceptable*. It is a *grateful* task to be the bearer of *welcome* intelligence to our friends.

I cannot but think the following letter from the Emperor of China to the Pope of Rome, proposing a coalition of the Chinese and Roman Churches, will be *acceptable* to the curious.

STALE.

The kids with pleasure browse the bushy plain; The shepherds are *grateful* to the swelling grain. DAYKIN.

Whatever is remote from common appearance is always *welcome* as vulgar as to childish crudity. JOUSSAUX.

#### ACCEPTANCE, ACCEPTATION,

Though both derived from the verb accept, have this difference, that the former is employed to express the abstract action generally; the latter only in regard to the single object of words. A book, or whatever else is offered to another, is *acceptable* in its own nature, but *accepted* only in its application to the acceptance of it.

*clasificadas, no de acuerdo con su sonido u ortografía, sino estrictamente de acuerdo con su significacion*<sup>6</sup>.

El diccionario se compone de un índice alfabético que recoge todas las posibles acepciones de una misma palabra, y de un índice que agrupa los vocablos por categorías semánticas, (figuras 17 y 18) organizándolas además en función de su morfología. La localización de un concepto en el *Thesaurus* se realiza en dos etapas. En primer lugar, se busca en el índice una palabra que describa determinada idea y se elige la entrada cuyo significado se aproxime más al concepto deseado. Por ejemplo, si se desea información acerca de la palabra *game*, en el índice alfabético aparece:

**game**

- n.* animal life 414.1
- cards 878.35
- contest 796.3
- fun 878.2
- influence 172.3
- intrigue 654.6
- laughingstock 967.7
- meat 308.12
- objective 653.2
- plan 654.1
- quarry 655.7
- sport 878.9
- stratagem 735.3
- vocation 656.6
- word list 878.34
- v.* gamble 515.18
- adj.* bold 893.18
- crippled 692.32
- plucky 624.14

---

<sup>6</sup>ROGET, P.M. *Roget's International Thesaurus*. 4ª ed. revised by Robert L. Chapman. New York [etc.]: Harper Collins Publishers, 1977, p. X.

matic, derivational; affixal, prefixal, infixal, suffixal.

### 583. NOMENCLATURE

- .1 NOUNS nomenclature, terminology, orismology, glossology [archaic]; onomatology, onomastics; toponymy, place-names, place-naming; autonomasia; polyonymy; taxonomy, classification, systematics, biosystematy, biosystematics, cytotaxonomy, binomial nomenclature, binomialism, Linnaean method, trinomialism; kingdom, phylum, class, order, family, genus, species.
- .2 naming, calling, denomination, appellation, designation, styling, terming, definition, identification; christening, baptism; dubbing; nicknaming.
- .3 name, onomato-, -onym, -onymy; appellation, appellative, denomination, designation, style, *nomen* [L.], cognomen, cognomination; proper name or noun; moniker or handle [both slang]; title, honorific; empty title or name; label, tag; epithet, byword; scientific name, trinomen, trinomial name, binomen, binomial name; *nomen nudum*, hyponym; tautonym; typonym; middle name; eponym; namesake; secret name, cryptonym, cuonym.
- .4 first name, forename, Christian name, given name, baptismal name.
- .5 surname, last name, family name, cognomen, byname; maiden name; married name; patronymic, matronymic.
- .6 (Latin terms) *praenomen*, *nomen*, *agnomen*, *cognomen*.
- .7 nickname, sobriquet, byname, cognomen; pet name, diminutive, hypocoristic, affectionate name.
- .8 alias, pseudonym, anonym, assumed name, false or fictitious name, *nom de guerre* [Fr]; pen name, *nom de plume*; stage name, *nom de theatre* [Fr], professional name; John Doe, Jane Doe, Richard Roe.
- .9 misnomer, wrong name.
- .10 signature, sign manual, autograph, hand, John Hancock [informal]; mark, mark of signature, cross, christcross, X; initials; subscription; countersignature, countersign, countermark, counterstamp; endorsement; visa, *visé* [Fr]; monogram, cipher, device; seal, sigil, signet.
- .11 VERBS name, denominate, nominate, designate, call, term, style, dub; specify; define, identify; title, entitle; label, tag; nickname; christen, baptize.
- .12 misname, misnomer, miscall, misterm, misdesignate.
- .13 be called, be known by or as, go by, go as, go by the name of, go or pass under the name of, bear the name of, rejoice in the name of; go under an assumed or false name, have an alias.
- .14 ADJS named, called, yecept [archaic], styled, titled, denominated, denominate [archaic], known as, known by the name of, designated, termed, dubbed, identified as; christened, baptized; what one may well or fairly or properly or fitly call.
- .15 nominal, cognominal; titular, in name only, nominative, formal; so-called, quasi; would-be, *soi-disant* [Fr]; self-called, self-styled, self-christened; honorific; epithetic(al); hypocoristic, diminutive; by name, by whatever name, under any other name.
- .16 denominative, nominative, appellative; eponymous, eponymic.
- .17 terminological, nomenclatural, orismological; onomastic; toponymic, toponymous; taxonomic, classificatory, binomial, Linnaean, trinomial.

### 584. ANONYMITY

- .1 NOUNS anonymity, anonymousness, namelessness; incognito; anonym.
- .2 what's-its-name or what's-his-name or what's-his-face or what's-her-name or what-you-may-call-it or what-you-may-call-em or what-d'ye-call-em or what-d'ye-call-it or whatzit [all informal]; *je ne sais quoi* [Fr], I don't know what; such-and-such; so-and-so, certain person, Mr. X; you-know-who.
- .3 ADJS anonymous, anon.; nameless, unnamed, unidentified, undesignated, unspecified, innominate, without a name, unknown; undefined; unacknowledged; incognito; cryptonymous, cryptonymic.

### 585. PHRASE

- .1 NOUNS phrase, expression, locution, utterance, usage, term, verbalism, -logy or -logia; word-group, construction, endocentric construction, headed group, syntagm; syntactic structure; noun phrase, verb phrase, verb complex, adverbial phrase, adjectival phrase, prepositional phrase; clause; sentence, period, periodic sentence; paragraph; idiom, idiotism, phrasal idiom; turn of phrase or expression, peculiar expression, manner or way of speaking; set phrase or term; conven-

- sound 450.1  
 television reception 345.5  
 unintelligibility 549.7  
 v. be noisy 453.8  
 sound 450.14  
 noise about 559.10  
 noiseless 451.10  
 noisemaker 453.5  
 noiseproof 159.18  
 noisily 453.13  
 noisome bad 675.9  
 n. harmful 675.12  
 malodorous 437.5  
 offensive 864.18  
 unhealthful 684.5  
 unsavory 429.7  
 noisy blustering 911.4  
 loud 453.12  
 vociferous 459.10  
 no laughing matter 672.3  
 nomad  
 n. Bedouin 274.4  
 adj. wandering 273.36  
 nomadism 273.3  
 no-man's-land 802.2  
 no matter 673.22  
 no matter what  
 by no means 524.9  
 whatever 79.6  
 nom de plume 583.8  
 nomen name 583.3  
 Roman name 583.6  
 nomenclature 583  
 nominal  
 n. noun 586.5  
 adj. cheap 849.7  
 grammatical 586.17  
 in name only 583.15  
 nominal charge 849.2  
 nominalism 80.1  
 nominally  
 cheaply 849.10  
 falsely 616.35  
 namely 80.18  
 nominate  
 appoint 780.10  
 choose 637.19  
 name 583.11  
 support politically 744.41  
 nominated 637.26  
 nomination  
 appointment 780.2  
 choice 637.8  
 consecration 1037.10  
 political nomination 744.11  
 nominative  
 n. case 586.9  
 adj. naming 583.16  
 nominal 583.15  
 nominee 780.8  
 no more dead 408.30  
 extinct 2.10  
 past 119.7  
 vanished 447.4  
 non(a)- 99.21  
 nonacceptance  
 refusal 776.1  
 rejection 638.1  
 nonaccomplishment  
 failure 725.1  
 nonachievement 723  
 nonadherent  
 incoherent 51.4  
 nonobservant 769.5  
 nonadmission 77.1  
 nonage  
 immaturity 124.3  
 nine 99.5  
 nonagenarian  
 ninety 99.7  
 old man 127.2  
 nonaggression 803.4  
 nonaggression pact 771.2  
 nonaggressive  
 indolent 708.18  
 unbelligerent 803.10  
 nonalcoholic 997.4  
 nonalcoholic beverage 308.48  
 nonaligned  
 independent 762.21  
 neutral 806.7  
 nonaligned nation-  
 country 181.1  
 independent 762.12  
 nonalignment 806.1  
 nonattendance 187.4  
 nonattendant 187.10  
 nonbeliever 1031.11  
 nonbelieving 503.8  
 nonbelligerent 803.5  
 nonbreakable 359.5  
 nonce, the 120.1  
 nonce word 582.8  
 nonchalance  
 apathy 856.4  
 inexcitability 858.5  
 unconcern 636.2  
 nonchalant  
 apathetic 856.13  
 unconcerned 636.7  
 unexcited 858.15  
 nonchalantly  
 calmly 858.18  
 unconcernedly 636.10  
 noncoherent 51.4  
 noncohesion 51  
 noncohesive  
 incoherent 51.4  
 separate 49.20  
 noncom 749.19  
 noncombatant  
 n. nonbelligerent 803.5  
 adj. unbelligerent 803.10  
 noncommissioned  
 officer  
 enlisted man 800.8  
 sergeant 749.19  
 supervisor 748.2  
 noncommittal 895.8  
 noncommitted 806.7  
 noncommunicable  
 indescribable 920.13  
 unimpartable 813.10  
 noncompetitive 786.5  
 noncompletion 723.1  
 noncompliance  
 disobedience 767.1  
 nonconformity 83.1  
 nonobservance 769.1  
 refusal 776.1  
 noncompliant 769.5  
 non compos mentis 473.25  
 nonconducting 342.33  
 nonconductor 342.13  
 nonconforming  
 disobedient 767.8  
 dissenting 522.6  
 nonobservant 769.5  
 uncompliant 83.5  
 nonconformist  
 n. dissenter 522.3  
 eccentric 474.3  
 heretic 1025.5  
 misfit 27.4  
 odd person 85.4  
 original 83.3  
 rebel 767.5  
 adj. individualistic 27.9  
 nonuniform 18.3  
 opposing 178.8  
 nonconformity  
 deviation 219.1  
 difference 16.1  
 disobedience 767.1  
 dissent 522.1  
 eccentricity 474.1  
 extraneousness 78.1  
 inconsistency 27.2  
 individuality 80.1  
 noncompliance 83  
 nonobservance 769.1  
 nonuniformity 18.1  
 perverseness 178.1  
 unorthodoxy 1025.1  
 noncontagious 813.10  
 noncontiguous 49.20  
 nonconvergent 218.6  
 noncooperation  
 disaccord 795.1  
 disobedience 767.1  
 opposition 790.2  
 resistance 792.1  
 noncooperative  
 insubordinate 767.9  
 opposing 790.8  
 resistant 792.5  
 noncreative 166.5  
 nondenominational  
 nonsectarian 1020.26  
 universal 79.14  
 nondescript  
 formless 247.4  
 simple 902.6  
 nondutiable 846.18  
 none  
 n. not a one 2.3  
 adv. not at all 2.11  
 nonelastic 356.12  
 nonemotional 856.9  
 nonentity  
 a nobody 673.7  
 mediocrity 680.5  
 nonexistence 2.1  
 thing of naught 4.2  
 weakening 160.6  
 none of your business 238.10  
 nonessential  
 n. accessory 6.2  
 adj. accessory 6.4  
 circumstantial 8.7  
 irrelevant 10.6  
 needless 669.10  
 superfluous 663.17  
 unimportant 673.15  
 nonesuch marvel 920.2  
 odd thing 85.5  
 the best 674.8  
 nonetheless  
 anyhow 657.12  
 notwithstanding 33.8  
 nonexistence  
 absence 187.1  
 nonsubsistence 2  
 nonexistent  
 absent 187.10  
 imaginary 535.19  
 nonsubsistent 2.7  
 vanished 447.4  
 nonexpectant 540.9  
 nonfactual 535.21  
 nonfeasance  
 mismanagement 734.6  
 neglect 534.1  
 nonaccomplishment 723.1  
 nonobservance 769.1  
 sin 982.2  
 nonfertile 166.4  
 nonfiction 602.10  
 nonflammability 332.1  
 nonflammable 332.9  
 nonfulfillment  
 insufficiency 662.1  
 nonaccomplishment 723.1  
 nonobservance 769.1  
 nonfunctional 669.14  
 nonhuman 85.15  
 nonimitation 23  
 noninfectious 813.10  
 nonintellectual 477.14  
 noninterference  
 freedom 762.9  
 neglect 534.1  
 political policy 744.4  
 noninterruption 135.2  
 nonintervention  
 avoidance 631.1  
 freedom 762.9  
 political policy 744.4  
 nonintoxicating 997.4  
 noninvolvement  
 avoidance 631.1  
 inaction 706.1  
 neutrality 806.1  
 non-few 1025.6  
 nonliving 382.5  
 nonliving matter 382.1  
 nonmalignant 674.20  
 nonmandatory 622.7  
 nonmedical therapist 688.12  
 nonmedical therapy 689.3  
 nonmetal 383.3

Si elegimos la acepción *sport*, pasamos en una segunda etapa a buscar, mediante el código que acompaña a dicha entrada en el índice alfabético, 878.9, en el índice clasificado del thesaurus el mencionado concepto, acompañado de todos sus posibles sinónimos y asociaciones de ideas. El código 878 corresponde al campo semántico *amusement*, que engloba al concepto deseado:

878.9 *game, sport* 878.34; *play; contest* 796.3;  
race 796.12; *event; meet; bout; match; go*  
[informal]; singles, doubles; twosome,  
threesome, foursome; double-header;  
pentathlon; play-off, runoff; games  
of chance 515.8.

El *Thesaurus* es fundamentalmente un útil para la transformación de ideas en palabras, principal objetivo con el que Roget lo concibiera, tal y como el propio título original de la obra expresa ... *para facilitar la expresión de ideas y servir de ayuda en la composición literaria*.

La ventaja de este tipo de diccionario sobre los ya existentes estriba en la posibilidad que ofrece de expresar diferentes aspectos de un concepto determinado, que, en definitiva, es el principio que determina el establecimiento de relaciones semánticas entre los elementos componentes de los tesauros documentales.

Wall considera que las siguientes características del Roget's Thesaurus:

1. *Conecta idea-palabras con palabras del texto;*
2. *Las idea-palabras están clasificadas;*
3. *Presenta en un índice alfabético todas las palabras que componen el thesaurus,*

se mantienen en el concepto de thesaurus para la recuperación de información:

1. *El thesaurus documental une palabras-clave o palabras codificadas con palabras del texto;*
2. *Las palabras clave están clasificadas;*
3. *Tiene un índice alfabético con todas las palabras que lo componen.*<sup>7</sup>

Estas características, a diferencia de los lenguajes documentales convencionales (listas de encabezamientos de materias, sistemas de clasificación) permiten fácilmente enviar al indizador y al usuario de las palabras del texto y las empleadas en la pregunta de recuperación, a las palabras clave.

La organización alfabética del Roget's, con sus referencias de *véase* que reenvían los términos específicos a términos genéricos, la diferenciación de sinónimos y la organización jerárquica de sus entradas, permiten un control del vocabulario que hace posible considerarlo no sólo precursor de los tesauros documentales sino que incluso podría ser utilizado en la actualidad como herramienta de control terminológico en recuperación de información.

En otros países se publicaron diccionarios inspirados en la filosofía del Roget's. Diez años después de la aparición de este thesaurus, en 1862, se publicó en Francia el *Dictionnaire analogique de la langue française*<sup>8</sup>, con sus mismas características, preparado por Boissière. En España se tardó más en realizar una obra parecida. Julio

---

<sup>7</sup>VICKERY, B.C. Thesaurus - a new word in documentation. *Journal of Documentation*, v. 16, n° 4 (1960) p. 187.

<sup>8</sup>BOISSIÈRE, P. *Dictionnaire analogique de la langue française, Répertoire complet des MOTS par les IDÉES et des IDÉES par les MOTS*. Paris: Aug. Boyer et Ce., Libraires-Éditeurs, 1862.

Casares fue el responsable de la elaboración del *Diccionario ideológico* que responde a los principios de organización conceptual del léxico. Dicho diccionario, cuya primera edición apareció en 1942, tiene como principal finalidad la de *poner a disposición del lector, mediante un inventario metódico, el inmenso caudal de voces castizas que, por desconocidas u olvidadas, no nos prestan servicio alguno; voces cuya existencia se sabe o se presupone, pero que, dispersas y como agazapadas en las columnas de los diccionarios corrientes, nos resultan inasequibles mientras no conozcamos de antemano su representación escrita*<sup>9</sup>.

Para alcanzar dicha finalidad se procede a una sistematización del vocabulario, reuniendo en grupos conceptualmente homogéneos cuantas palabras guardan relación con una idea determinada. *El lector que examine cualquiera de estos grupos no sólo se dará cuenta fácilmente del alcance de este trabajo, sino que quedará sorprendido al ver la enorme riqueza de medios expresivos a cuyo uso nunca pudo tener acceso. (Casares: 1977, p. XIII)*<sup>10</sup>.

La parte nuclear del diccionario consiste en la organización del léxico en una serie de grupos de afinidad semántica, ordenados alfabéticamente por la palabra que les sirve de *enunciado* o *cabeza*. Esta disposición es la que le da la capacidad de inducción, característica que comparte con los tesauros documentales. A dicha característica se refiere el autor cuando escribe acerca de la necesidad de *crear, junto al actual registro por abecé, archivo hermético y desarticulado, el diccionario orgánico, viviente, sugeridor de imágenes y asociaciones, donde, al conjuro de la idea, se ofrezcan en tropel las voces, seguidas del utilísimo cortejo de sinonimias, analogías, antítesis y referencias; un*

---

<sup>9</sup>CASARES, J. *Diccionario ideológico de la lengua española: Desde la idea a la palabra; desde la palabra a la idea*. 2ª ed. Barcelona: Gustavo Gili, 1977, p. XIII.

<sup>10</sup>CASARES, J. opus cit., p. XIII.

*diccionario comparable a esos bibliotecarios solícitos que, poniendo a contribución el índice de materias, abren camino al lector más desorientado, le muestran perspectivas infinitas y le alumbran fuentes de información inagotables...*<sup>11</sup>

Actualmente se sigue utilizando el vocablo en latín y en castellano para referirse a diccionarios lingüísticos, así como de sinónimos e ideas afines, tanto en España como en el extranjero<sup>12</sup>.

### 3.2. Tesoros documentales

Aunque un tesoro<sup>13</sup> como el de Roget y un tesoro documental no siguen los mismos propósitos<sup>14</sup>, ambos confluyen en un objetivo: evitar al usuario tener que pensar en todas las palabras que puedan expresar una determinada idea. El Roget enfoca esta ayuda al campo de la creación literaria, y los tesoros documentales, hacia el ahorro de tiempo y ayuda en el proceso de recuperación de información, ya que las palabras con

---

<sup>11</sup>CASARES, J. Nuevo Concepto de Diccionario de la Lengua, y otros problemas de Lexicografía y Gramática. En: **Obras Completas**. Madrid: Espasa Calpe, 1941, tomo V, p. 118.

<sup>12</sup>Por ejemplo, en ediciones recientes de estas obras: **Mega thesaurus: Sinónimos, antónimos, parónimos e ideas afines**. Barcelona: Sopena, 1994.

<sup>13</sup>Aunque el término *thesaurus* es de uso universal, nosotros optamos aquí por la denominación española *tesauro*, que es la que se emplea de forma generalizada para referirse a los tesoros aplicados a la documentación desde la publicación de la norma española UNE 50-106-1990.

<sup>14</sup>A. GARCIA GUTIÉRREZ delimita las diferencias entre tesoros lingüísticos y documentales en el trabajo: Los lenguajes documentales. En: **Fundamentos de información y documentación**, comp. López Yepes, J. Madrid: Eudema, 1990, p. 327: [...] un tesoro no se corresponde con lo que entendemos por diccionario estrictamente, a pesar de sus semejanzas, por dos motivos: 1º Los diccionarios se elaboran gracias a unos procedimientos provenientes de la Lexicografía, y no es objetivo de esa disciplina el establecer relaciones de diversa índole entre los términos que forman parte de su estudio. [...] Un tesoro no define los términos que recoge, excepto en las ocasiones en que existe ambigüedad y con fines puramente delimitatorios y prácticos. 2º A nivel estructural, un tesoro tiene múltiples formas de representación por lo que, en una fórmula definitiva general no puede concebirse como diccionario, lista, etc. Los índices anexos a los tesoros pueden tener mayor relación pero la existencia de descriptores compuestos elimina toda similitud.

significados similares son potencialmente sustituibles en la búsqueda.

Históricamente, el concepto de tesoro en relación con el almacenamiento y la recuperación de información se desarrolló como consecuencia de la *última revolución industrial... que nos inundó de todo tipo de documentos*<sup>15</sup> y con la aplicación de la automatización a dichos procesos (aunque un centro de documentación puede utilizarlo sin necesidad de mecanización, permite su integración en un sistema automatizado). *El thesaurus corresponde a la verdadera explosión de información y la revolución de la informática. Su organización y su estructura reflejan la complejidad de los datos a tratar y sus interrelaciones*<sup>16</sup>.

*Corrían los años finales de la década de los 40 cuando se empezó a pensar que las claves para determinar y clasificar [los documentos] debían ser precisamente las propias palabras con que se expresaban. Surgieron así las palabras clave, piedra angular de un nuevo sistema de clasificación*<sup>17</sup>.

*En 1947 Mooers, con cierto grado de confidencialidad, hacía referencia al uso de un thesaurus en asociación con un sistema de recuperación mecanizado*<sup>18</sup>, y en 1951 Taube sentó las bases de la postcoordinación en recuperación de información con el desarrollo del sistema uniterm. *Como un primer paso, se adoptó como palabra clave una*

---

<sup>15</sup>CURRÁS, E. **Tratado sobre Ciencia de la Información**. Rosario (Argentina): Universidad Nacional, 1996, p. 161.

<sup>16</sup>GUINCHAT & SKOURI. **Guide pratique des techniques documentaires**. 2 v. Paris: EDICEF, 1989, p. 127.

<sup>17</sup>CURRÁS, E. Op. cit., p. 162.

<sup>18</sup>ROBERTS, N. The pre-history of the information retrieval thesaurus. **Journal of Documentation**, v. 40, n° 4 (1984) pp. 271-285.

*sola palabra. Eso fue el origen del Sistema Uniterm... Casi simultáneamente en el tiempo se construyó el sistema Zatocoding, en el que las palabras claves designaban conceptos y éstos se podían expresar por una o más palabras*<sup>19</sup>. Wall<sup>20</sup> atribuye a Luhn el origen del tesoro utilizado en el campo de la documentación, el cual hizo pública, antes que nadie, la idea de que un tesoro podría dar la posibilidad a una máquina de pasar las palabras formuladas en las preguntas a palabras clave.

En relación con los mencionados descubrimientos hechos por Taube, Mooers y Luhn, en el campo de la recuperación de información, Izquierdo Arroyo y Moreno Fernández escriben:

*Estas tres personalidades, además de revolucionar las técnicas de indización, pusieron los jalones teóricos y metodológicos que permitirían la creación y uso de los thesauri en la década de los 60. El nacimiento del fichero invertido -Taube-, gracias al cual es posible acceder a la información de manera directa y no secuencial, la combinación de términos en la búsqueda, la indización por conceptos y el empleo de índices permutados, son otros tantos pilares sobre los que se orientan los modernos sistemas de documentación y los thesauri.*<sup>21</sup>.

Luhn sugirió la idea de elaborar un diccionario de nociones, ésto es, de agrupar palabras de parecido significado en familias de conceptos. Aplicó esta idea en 1951 utilizando un ordenador para registrar listas de palabras, provistas de referencias cruzadas,

---

<sup>19</sup>CURRÁS, E. Op. cit., p. 162.

<sup>20</sup>WALL, E. Information systems. **Chemical engineering progress**, v. 55 (1959) pp. 55-59, cit. por B. C. VICKERY: Thesaurus-A new word in documentation. **Journal of Documentation**, v. 6, nº 4 (1960) p. 185.

<sup>21</sup>IZQUIERDO ARROYO, J.M. & MORENO FERNÁNDEZ, L.M. Listas de encabezamientos de materia y Thesauri en perspectiva comparada. **Documentación de las Ciencias de la Infomación**. nº 17 (1994) p. 292.

que después empleaba para indizar automáticamente por materias. De la fuente de inspiración de este procedimiento dió idea el propio Luhn cuando escribió en 1957: *Ésto es tan similar al trabajo requerido para la creación de un thesaurus como el Roget's, que la organización básica de tal obra parece servir como esqueleto para este proceso... El resultado físico puede ser un diccionario en dos partes. La primera parte sería una lista presentada en algún orden sistemático de familias de nociones, cada una de ellas identificada con un símbolo tal como un número o una palabra clave. La segunda parte sería un índice alfabético de las palabras del texto que aparecen en la primera parte, dando la palabra clave de una o varias familias de nociones de las cuales la palabra dada [del texto] es miembro*<sup>22</sup>.

También en 1957 Bernier y Heumann<sup>23</sup> propusieron el uso de un tesoro para organizar el vocabulario en sistemas de recuperación de información. La propuesta tenía como principal objetivo ayudar al usuario en el proceso de búsqueda, quien podía, mediante el tesoro, conocer qué palabras clave utilizadas en un sistema respondían a sus necesidades particulares. Bernier y Heumann concebían el tesoro como una compleja red de relaciones entre palabras, con una complicada estructura de referencias cruzadas. Ellos fueron los primeros en introducir el concepto de término preferente y término no preferente.

Según Vickery<sup>24</sup>, la primera vez que se hizo referencia públicamente a la palabra thesaurus en conexión con la recuperación de información fue el 14 de Mayo de 1957 en

---

<sup>22</sup>LUHN, H.P. Auto-encoding of documents for information retrieval systems. **Modern trends in documentation**. Cambridge: Pergamon Press, 1959, pp. 45-58.

<sup>23</sup>BERNIER, C.L.; HEUMANN, K.F. Correlative indexes III: semantic relations among sermanterms -the technical thesaurus. **American Documentation**, n ° 8 (1957) pp. 211-220.

<sup>24</sup>VICKERY, B.C. Op. cit., p. 181.

una comunicación presentada por Helen Brownson en la *Dorking conference on classification*, en donde sostiene que *el problema [de la recuperación de información], como algunos investigadores han visto por fin, es transformar los conceptos y sus relaciones, que aparecen expresados en el lenguaje de los documentos, en algún lenguaje más normalizado, con sinónimos controlados y estructuras sintácticas simplificadas... Cuanto más avancemos en la mecanización de las técnicas de traducción del lenguaje normal en un lenguaje manejable mediante ordenador, más probabilidades tendremos de poder realizar búsquedas automáticas de información a gran escala... [Algunos investigadores] han ideado que la mejor solución puede ser la aplicación de un tesoro automatizado basado en redes de términos relacionados de mayor o menor amplitud.*<sup>25</sup>

Las mencionadas innovaciones incorporadas al campo de la recuperación de información: sistemas uniterm y zatocoding, alternativas propuestas por Luhn<sup>26</sup>, dieron como resultado el nacimiento de los primeros lenguajes postcoordinados, que empezaron a construirse entre finales de los años 50, y primeros de los 60. El primer tesoro documental del que se tiene constancia data de 1959, y fue elaborado por la organización Du Pont, aunque tuvo escasa difusión. Las fechas comprendidas entre 1960 y 1967 fueron prolíficas en el desarrollo de estos lenguajes documentales, especialmente en los Estados Unidos. En 1960 se publicó el *Thesaurus of ASTIA Descriptors*<sup>27</sup>, de Armed Services Technical Information Agency, Arlington Va, USA. Se creó con la idea de utilizarlo en sistemas de recuperación automatizados y consistía en una simple lista de términos en orden alfabético, con pocas referencias cruzadas. En 1961 el American

---

<sup>25</sup>BROWNSON, H. *Proceedings of the international study conference on classification for information retrieval*. London: Aslib, 1957, pp. 99-100.

<sup>26</sup>LUHN, H.P. A new method of recording and searching information. *American Documentation*, v. 4, n° 1 (1953) pp. 14-16.

<sup>27</sup>THESAURUS of ASTIA descriptors. Arlington (Virginia): Armed Services Technical Information Agency [Government Printing Office], 1960.

Institute of Chemical Engineers auspició el *Chemical Engineers Thesaurus*<sup>28</sup>. En el seno del Engineers Joint Council se llevó a cabo la preparación, en 1964, del *Thesaurus of Engineers Terms*<sup>29</sup>, cuya segunda edición constituiría uno de los más amplios y completos tesauros de la historia: el *Thesaurus of Engineers and Scientific Terms (TEST)*, editado en 1967 por el Departamento de Defensa de los Estados Unidos y la Junta de Ingenieros de Nueva York, que contiene 23.364 términos en sus 890 páginas.

### 3.2.1. Formas alternativas de organización de los tesauros

Como ya se ha dicho, es posible organizar un tesoro de diferentes formas. Por lo común, este lenguaje documental presenta los términos estructurados alfabética y sistemáticamente, pero, a lo largo de su historia, se han puesto en práctica formas de organización alternativas. Lancaster<sup>30</sup> analiza el primer intento en este sentido, realizado por Barhydt y Schmidt<sup>31</sup> en 1968, quienes desarrollaron un tesoro de educación, que se componía de dos partes complementarias y al propio tiempo independientes: un sistema de clasificación con sus relaciones jerárquicas, y un tesoro alfabético dotado de relaciones semánticas interjerárquicas, que contribuían a indizar y recuperar la información con eficacia.

---

<sup>28</sup>AMERICAN INSTITUTE OF CHEMICAL ENGINEERS. *Chemical engineering thesaurus; a wordbook for use with the concept coordination system of information storage and retrieval*. New York: American Institute of Chemical Engineering, 1961.

<sup>29</sup>THESAURUS of Engineers Terms. New York: Engineers Joint Council, 1961.

<sup>30</sup>LANCASTER, F.W. Vocabulary control in information retrieval systems. *Advances in librarianship*, n° 7 (1977) p. 13.

<sup>31</sup>BARHYDT, G.T. & SCHMIDT, C.T. *Information Retrieval Thesaurus of Education Terms*. Cleveland (Ohio): Case Western Reserve University, 1968.

Pero la más famosa alternativa de estas características fue el *Thesaurofacet*<sup>32</sup>, que integra también un sistema de clasificación y un tesoro. De igual modo, la English Electric Company organizaba su información por medio de un sistema de clasificación por facetas, la *EE Classification for Engineering*; durante la compilación de su cuarta edición, a finales de los años 60, se estimó oportuno incorporar al sistema algún procedimiento que permitiera la indización postcoordinada y el uso de los avances informáticos. Resultado de las investigaciones realizadas con este objetivo fue la creación del *Thesaurofacet*, compuesto por un sistema de clasificación por facetas y por un tesoro, de manera que los términos aparecen en ambos lenguajes y son complementarios.

El *Thesaurofacet* tiene dos características particulares. En primer lugar, la organización de la información en facetas posibilita su adaptación a distintos tipos de sistemas de recuperación de información. En segundo lugar, a las ventajas de un tesoro se suman las de un sistema de clasificación, derivando en una mayor capacidad para establecer relaciones entre conceptos que las que pueden conseguirse con el uso independiente de cada uno de los componentes de este peculiar lenguaje documental.

Para utilizar el *Thesaurofacet*, el usuario busca un término en el tesoro, en donde encuentra un número de clasificación, que viene identificado por una notación numérica, al tiempo que le aporta información sobre cualquier posible término relacionado (al igual que en cualquier tesoro). La clasificación organiza la información en campos temáticos agrupados según la distribución convencional en ciencia y tecnología. Dentro de cada campo de materias los términos se disponen de acuerdo con los principios de la clasifi-

---

<sup>32</sup>AITCHISON, J. The Thesaurofacet: a multipurpose retrieval language tool. *Journal of Documentation*, v. 26, (3) 1970, p. 187.

cación por facetas, lo que significa que dentro de cada sección, los conceptos se distribuyen en grupos homogéneos de acuerdo con características bien definidas.

La disposición gráfica del vocabulario es otra importante variación en cuanto a forma de presentación. Los primeros ejemplos de tesauros organizados en dos dimensiones fueron el *Circular thesaurus* desarrollado por Netherlands Armed Service Technical Information Agency (TDCK) en 1962 y el *Thesaurus EURATOM*. Esta forma de presentación ha facilitado la elaboración de tesauros multilingües y resulta muy cómoda de consultar, porque cada gráfico que lo compone corresponde a un único campo semántico.

A partir de 1960 se ha producido una enorme cantidad de tesauros para la recuperación de información. La mayoría en lengua inglesa y destinados a las áreas de la ciencia y tecnología. En menor cantidad existen tesauros en campos no científicos y en otras lenguas, habiéndose construido, además, tesauros multilingües.

### **3.3. Desarrollo de tesauros documentales en España**

#### **3.3.1. Introducción**

Hasta la aparición de los primeros tesauros, se manejaban en España vocabularios desarrollados por instituciones nacionales e internacionales, concienciadas de los inconvenientes que genera el uso de clasificaciones de carácter universal para la recuperación de información especializada. Estos vocabularios, por lo general, carecían de una estructura, salvo la ordenación alfabética.

En un principio, el desarrollo de lenguajes postcoordinados estuvo limitado al Consejo Superior de Investigaciones Científicas (CSIC), en cuyo seno quedaba restringido su uso. En las bibliotecas se utilizaban lenguajes precoordinados (clasificaciones y listas de encabezamientos de materias) y se puede decir que la situación en la actualidad no ha variado en la medida en la que lo ha hecho en otros países. La elaboración de tesauros en España es escasa, prueba de ello es que el censo de los publicados en este momento, que incluimos en el epígrafe 3.4. de este mismo capítulo, es de 30. Bien es cierto que existen numerosos vocabularios controlados en algunos centros de documentación, por ejemplo listas de descriptores o de palabras clave, que se presentan bajo la denominación de tesauros, y que no registramos en el censo porque se limitan a listar términos, sin estructurar y sin establecer ningún tipo de relación semántica entre ellos. Entre otros, podríamos mencionar el denominado *Leda thesaurus*<sup>33</sup> (figura 19) compuesto por un total de 3.674 lexemas sobre legislación educativa de carácter general.

### 3.3.2. Los tesauros en el ámbito formativo

La información precedente es fácilmente contrastable si se consideran los primeros planes de estudios de la Diplomatura en Biblioteconomía y Documentación (figuras 20, 21 y 22), en los que los tesauros se trataban de forma marginal, a diferencia de la importancia que se otorgaba a las clasificaciones.

Si nos remontamos a fechas anteriores a la implantación de la mencionada diplomatura, encontramos un vacío en lo que a lenguajes postcoordinados se refiere, tanto en la programación de la Escuela de Bibliotecarias de Barcelona, como en los estudios

---

<sup>33</sup>ESPAÑA. Ministerio de Educación y Ciencia. *LEDA Thesaurus*. [inédito], 1982.

*DISMINUIDO	T1	DE
*DISMINUIDO FISICO	T1	DE
*DISMINUIDO PSQUICO	T1	DE
*DISMINUIDO SENSORIAL	T1	DE
*DISPENSA ESCOLARIDAD	T1	DE
*DISPENSA TITULO	T1	DE
*DISPOSICION	T1	DE
*DISTRIBUCION ALUMNADO	T1	DE
*DISTRIBUCION CENTROS	T1	DE
*DISTRIBUCION TASA JUEGO	T1	DE
*DISTRITO UNIVERSITARIO	T1	DE
*DOCTOR	T1	DE
*DOCTOR ARQUITECTO	T1	DE
*DOCTOR BELLAS ARTES	T1	DE
*DOCTOR CIENCIAS	T1	DE
*DOCTOR CIENCIAS BIOLOGICAS	T1	DE
*DOCTOR CIENCIAS ECONOMICAS EMPRESARIALES	T1	DE
*DOCTOR CIENCIAS FISICAS	T1	DE
*DOCTOR CIENCIAS GEOLOGICAS	T1	DE
*DOCTOR CIENCIAS INFORMACION	T1	DE
*DOCTOR CIENCIAS MAR	T1	DE
*DOCTOR CIENCIAS MATEMATICAS	T1	DE
*DOCTOR CIENCIAS POLITICAS SOCIOLOGIA	T1	DE
*DOCTOR CIENCIAS QUIMICAS	T1	DE
*DOCTOR DERECHO	T1	DE
*DOCTOR FARMACIA	T1	DE
*DOCTOR FILOLOGIA	T1	DE
*DOCTOR FILOSOFIA CIENCIAS EDUCACION	T1	DE
*DOCTOR FILOSOFIA LETRAS	T1	DE
*DOCTOR GEOGRAFIA HISTORIA	T1	DE
*DOCTOR INFORMATICA	T1	DE
*DOCTOR INGENIERO AERONAUTICO	T1	DE
*DOCTOR INGENIERO AGRONOMO	T1	DE
*DOCTOR INGENIERO CAMINOS	T1	DE
*DOCTOR INGENIERO INDUSTRIAL	T1	DE
*DOCTOR INGENIERO MINAS	T1	DE
*DOCTOR INGENIERO MONTES	T1	DE
*DOCTOR INGENIERO NAVAL	T1	DE
*DOCTOR INGENIERO TELECOMUNICACIONES	T1	DE
*DOCTOR MEDICINA	T1	DE
*DOCTOR MEDICINA CIRUGIA	T1	DE
*DOCTOR PSICOLOGIA	T1	DE
*DOCTOR VETERINARIA	T1	DE
*DOCTORADO	T1	DE
*DOCUMENTACION	T1	DE
*DROGODEPENDENCIA	T1	DE
*EATP	T1	DE
*ECUADOR	T1	DE
*EDAD MAXIMA ALUMNADO	T1	DE
*EDAD MINIMA ALUMNADO	T1	DE
*EDAD REGLAMENTARIA ALUMNADO	T1	DE
*EDIFICIO ESCOLAR	T1	DE
*EDUCACION A DISTANCIA	T1	DE
*EDUCACION AMBIENTAL	T1	DE
*EDUCACION CIVICO SOCIAL POLITICA	T1	DE
*EDUCACION COMPENSATORIA	T1	DE
*EDUCACION EMPLEO	T1	DE
*EDUCACION EN EL EXTRANJERO	T1	DE
*EDUCACION ESPECIAL	T1	DE
*EDUCACION FISICA	T1	DE
*EDUCACION INFANTIL	T1	DE
*EDUCACION MUSICAL	T1	DE

32903 ORDEN de 27 de octubre de 1983 por la que se aprueba el plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de Granada, dependiente de la Universidad de Granada.

El Sr. Visto la propuesta formulada por el Rectorado de la Universidad de Granada en solicitud de aprobación del plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de Granada dependiente de la Universidad de Granada.

Considerando que la citada propuesta ha sido informada favorablemente por la Junta de gobierno de la Universidad de Granada y de acuerdo con lo establecido en la Ley 14/1970, de 4 de agosto, General de Educación y Financiamiento de la Reforma Educativa, la Orden ministerial de 24 de febrero de 1981 (Boletín Oficial del Estado de 14 de marzo), por la que se establecieron las directrices para la elaboración de los planes de estudio de las Escuelas Universitarias de Biblioteconomía y Documentación, y visto el dictamen favorable emitido por la Junta Nacional de Universidades.

Este Ministerio ha dispuesto:

Primero.—Queda aprobado, con efectos desde la fecha de su implantación, el plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de Granada, dependiente de la Universidad de Granada, quedando estructurado el plan de estudios de la forma que se indica en anexo adjunto a esta Orden ministerial.

Segundo.—La aprobación de este plan de estudios no implica aumento del gasto público.

Lo que comunico a V. I. para su conocimiento y efectos. Madrid, 27 de octubre de 1983.—P. D., el Director general de Enseñanza Universitaria, Emilio Lamo de Espinosa.

Ilmo. Sr. Director general de Enseñanza Universitaria.

ANEXO QUE SE CITA

Plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de la Universidad de Granada que fue creado por Real Decreto 1618/1982, de 18 de junio (Boletín Oficial del Estado de 23 de julio)

PRIMER CURSO

Objetivos: Introducción a la comunicación, al conocimiento del documento y a las bases de las Ciencias Bibliotecarias y Documentarias. Adquisición de la terminología fundamental.

Materias comunes	Horas semanales de clase	
	Teóricas	Prácticas
Historia del documento y de los depósitos documentales	3	El equivalente a un mínimo de cinco horas semanales
Biblioteconomía I (organización de bibliotecas)	3	
Bibliografía I (Introducción a las fuentes de información)	3	
Análisis documental I (principios generales de catalogación)	3	
Idioma moderno	3	
Seminario de estudios I (cuatrimestral)		
Materias optativas de carácter monográfico a determinar	3	
Seminario de estudios II (cuatrimestral)		
Materias optativas de carácter monográfico a determinar	3	

SEGUNDO CURSO

Objetivos: Estudio, teoría y aplicación de las técnicas bibliotecarias y documentarias, de las normas de catalogación y clasificación y uso de las fuentes de información bibliográfica.

Materias comunes	Horas semanales de clase	
	Teóricas	Prácticas
Biblioteconomía II (teoría, historia y biblioteconomía especializada)	3	El equivalente a un mínimo de cinco horas semanales
Bibliografía II (bibliografía general y especializada)	3	
Documentación (teoría, historia y organización de Centros)	3	
Análisis documental II (catalogación)	3	
Lenguajes documentarios I (principios generales de la clasificación)	3	
Idioma moderno	2	

TERCER CURSO

Objetivos: Inicio de especialización en las disciplinas bibliotecarias y bibliográficas, así como en el tratamiento de datos. Prácticas permanentes en bibliotecas y Centros de documentación metodológica del trabajo científico.

Materias comunes	Horas semanales de clase	
	Teóricas	Prácticas
Lenguajes documentarios II (sistemas de clasificación y análisis de contenido)	3	—
Información documental (automatización de bibliotecas y telecomunicación)	3	
Teoría y sociología de la comunicación	3	
Archivística	3	
Idioma moderno	2	
Seminario de estudios I (primer cuatrimestral)	3	

Prácticas en bibliotecas y Centros de documentación durante cuatrimestral.

Memoria al final de los estudios Superada la totalidad de las asignaturas que comprende este plan de estudios, los alumnos deberán elaborar una Memoria sobre un tema relacionado con las materias cursadas que será requisito previo para la obtención del título de Diplomado en Biblioteconomía y Documentación.

MINISTERIO DE EDUCACION Y CIENCIA

8620 **ORDEN** de 20 de marzo de 1984 por la que se aprueba el plan de estudios de la Escuela Universitaria (no estatal) de Biblioteconomía y Documentación de Barcelona, dependiente de la Universidad de Barcelona.

Ilmo. Sr.: Vista la propuesta formulada por el Rectorado de la Universidad de Barcelona en solicitud de aprobación del plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de Barcelona, dependiente de la mencionada Universidad;

Considerando que la citada propuesta ha sido informada favorablemente por la Junta de Gobierno de la Universidad de Barcelona, y de acuerdo con lo establecido en la Ley 14/1970, de 4 de agosto, General de Educación y Financiamento de la Reforma Educativa; la Orden de 24 de febrero de 1981 (Boletín Oficial del Estado, del 14 de marzo), por la que se establecieron las directrices para la elaboración de los planes de estudio de las Escuelas Universitarias de Biblioteconomía y Documentación, y visto el dictamen favorable emitido por la Junta Nacional de Universidades con fecha 10 de febrero de 1984.

Este Ministerio ha dispuesto:

Primero.—Se aprueba el plan de estudios de la Escuela Universitaria (no estatal) de Biblioteconomía y Documentación de Barcelona, adscrita a la Universidad de Barcelona, quedando estructurado el mismo de la forma que se indica en anexo adjunto a esta Orden.

Segundo.—Las asignaturas optativas se impartirán cuando haya un mínimo de 15 alumnos matriculados en cada asignatura.

Tercero.—La aprobación de este plan de estudios no implica aumento del gasto público.

Lo que comunico a V. I. para su conocimiento y efectos. Dios guarde a V. I. muchos años.

Madrid, 20 de marzo de 1984.—P. D., el Director general de Enseñanza Universitaria (Orden de 27 de marzo de 1982), Emilio Lamo de Espinosa.

Ilmo. Sr. Director general de Enseñanza Universitaria.

ANEXO QUE SE CITA

Plan de estudios de la Escuela Universitaria (no estatal) de Biblioteconomía y Documentación de Barcelona, adscrita a la Universidad de Barcelona.

Primer curso

Objetivos: Introducción a la comunicación, al conocimiento del documento y a las bases de las ciencias bibliotecarias y documentales. Adquisición de una terminología fundamental.

	Horas semanales
<b>Materias anuales obligatorias:</b>	
Historia del libro y de las bibliotecas	2
Biblioteconomía I (organización de bibliotecas)	3
Bibliografía I (introducción a las fuentes de la información)	3
Análisis documental I (principios generales de catalogación)	4
Idiomas modernos	2

	Horas semanales
<b>Materias cuatrimestrales obligatorias:</b>	
Terminología científica	2
Técnica del trabajo científico	2
Literatura	2
Historia	2
<b>Total</b>	<b>11</b>

Segundo curso

Objetivos: Estudio, teoría y aplicación de las técnicas bibliotecarias y documentales, de las normas de catalogación y de clasificación y del uso de las fuentes de información bibliográfica.

	Horas semanales
<b>Materias anuales obligatorias:</b>	
Biblioteconomía II (teoría, historia y biblioteconomía especializada)	3
Bibliografía II (bibliografía general y especializada)	3
Documentación I (teoría, historia y organización de centros)	2
Análisis documental II (catalogación)	4
Clasificación e indización	4
Idiomas modernos	2
Libro infantil	1

	Horas semanales
<b>Materias cuatrimestrales opcionales:</b> (Dos materias como mínimo hasta completar las 23 horas semanales):	
Artes Gráficas	2
Historia	2
Latin	2
Literatura catalana	2
Paleografía	2
Patología del papel	2
Principios de administración	2
Psicosociología del lector	2
<b>Total</b>	<b>23</b>

Tercer curso

Objetivos: Inicio de una especialización en las ciencias bibliotecarias y bibliográficas y en el tratamiento de datos. Prácticas permanentes en bibliotecas y en centros de documentación.

Uno de los dos cuatrimestres del tercer curso se dedicará a prácticas en bibliotecas o centros de documentación. Durante el otro cuatrimestre el alumno deberá cursar las siguientes materias:

	Horas semanales
<b>Materias cuatrimestrales obligatorias:</b>	
Informática y automatización de bibliotecas	4
Documentación II	4
Restauración y encuadernación	3
Archivística	2

	Horas semanales
<b>Materias cuatrimestrales opcionales (a elegir dos):</b>	
Catalogación (curso monográfico I)	2
Catalogación (curso monográfico II)	2
Lenguajes documentales (curso monográfico I)	3
Lenguajes documentales (curso monográfico II)	3
Bibliografía (curso monográfico)	2

	Horas semanales	Horas semanales
<b>Materias cuatrimestrales optativas (a elegir dos entre las siguientes o las no cursadas en segundo):</b>		
Banco de datos	2	
Bibliometría	2	
Estadística	2	4
Hemerotecas	2	
Material audiovisual	2	
<b>Total</b>	<b>20</b>	

después el 0,3 por 100 sobre dicho montante anual, y multiplicando ese resultado por los siete meses a los que hace referencia la revisión y dividiendo por los doce meses del año.

Y no habiendo más asuntos que tratar, siendo las diecisiete horas del día citado, se da por terminada la reunión, de la que se extiende la presente acta que, en prueba de conformidad, es suscrita por la totalidad de los asistentes.

**2019** Acuerdo 17 junio 1987 (Consejo de Universidades). ESCUELA UNIVERSITARIA DE BIBLIOTECONOMIA Y DOCUMENTACION. Plan de estudios de la de Salamanca.

Resuelve:

Homologar el plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de Salamanca, que quedará estructurado conforme figura en el anexo.

**ANEXO QUE SE CITA**

Plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de la Universidad de Salamanca

	Horas semanales	
	Teóricas	Prácticas
<i>Primer curso</i>		
Historia del Documento y de los Depósitos documentales . . . . .	3	1
Bibliografía . . . . .	3	1
Análisis Documental . . . . .	3	1
Español . . . . .	3	1
Idioma Moderno . . . . .	3	1
Dos asignaturas cuatrimestrales, a elegir entre:		
Latín.		
Paleografía Latina.		
Historia e Historiografía.		
Crítica y Comentario de Fuentes.		
Las demás optativas que ofrezca la Universidad cada año.		
Cada una de estas asignaturas optativas constará de cuatro horas semanales de clases (entre teóricas y prácticas) distribuidas según la naturaleza de las asignaturas.		
<i>Segundo curso</i>		
Documentación . . . . .	3	1
Lenguajes documentales: Principios generales de la clasificación . . . . .	3	1
Iniciación a la Informática . . . . .	3	1
Idioma Moderno . . . . .	3	1
Una de estas dos opciones, a elegir por el alumno:		
Opción de Biblioteconomía y Documentación:		
Biblioteconomía . . . . .	3	1
Opción de Archivística:		
Archivística . . . . .	3	1
Dos asignaturas cuatrimestrales, a elegir entre:		
Artes Gráficas.		
Paleografía Castellana.		

3974

Historia de las Instituciones del Antiguo Régimen.  
Conservación del Material Documental.  
Las demás optativas que ofrezca la Universidad cada año.  
Cada una de estas asignaturas optativas tendrá cuatro horas semanales, entre teóricas y prácticas, según su naturaleza.

**Tercer curso**

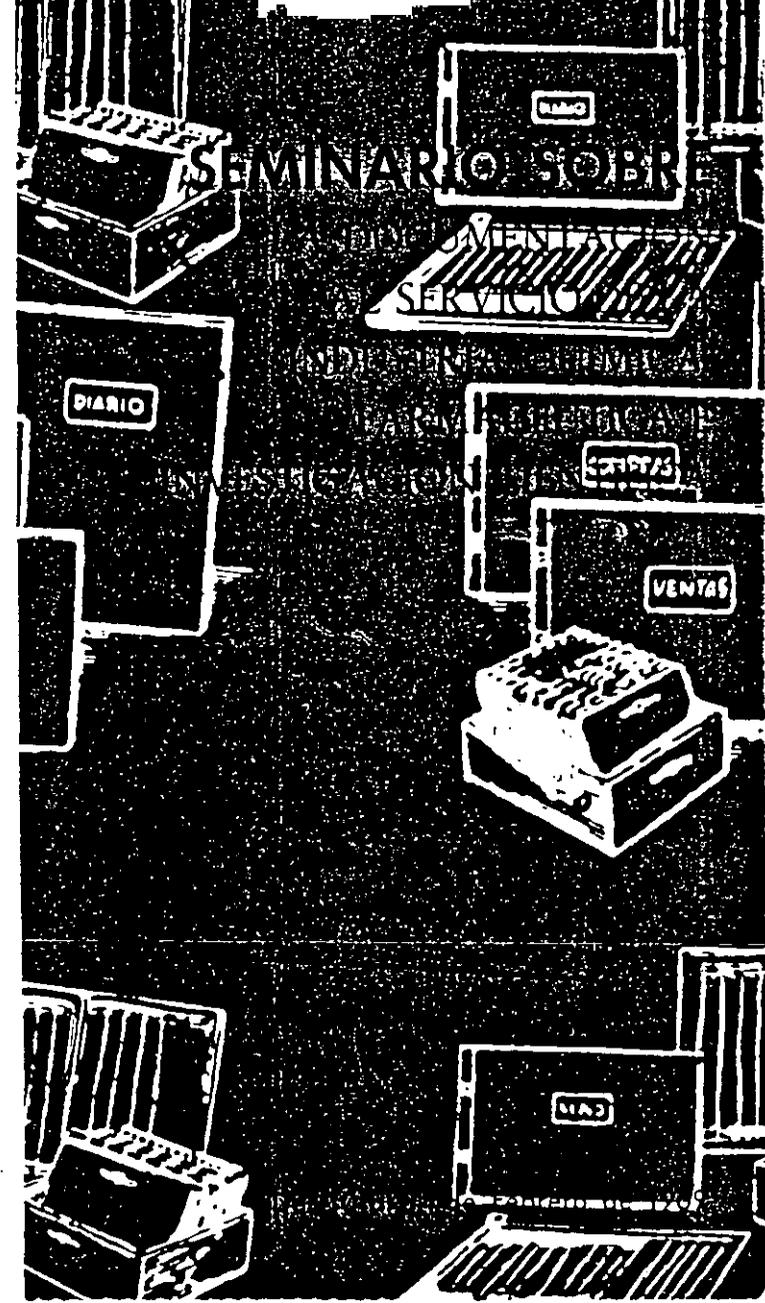
	Horas semanales
Informática Aplicada y Banco de Datos . . . . .	3
Idioma Moderno . . . . .	3
Una de estas opciones, a elegir por el alumno:	
Opción de Biblioteconomía y Documentación:	
Lenguajes Documentales de Bibliotecas . . . . .	3
Archivística para Bibliotecarios . . . . .	3
Opción de Archivística:	
Lenguajes Documentales de Archivos . . . . .	3
Biblioteconomía para Archiveros . . . . .	3
Dos asignaturas cuatrimestrales, a elegir entre:	
Restauración y Encuadernación.	
Diplomática.	
Técnicas de Trabajo Científico.	
Historia de la Administración Contemporánea.	
Las demás optativas que ofrezca la Universidad cada año.	
Cada una de estas asignaturas optativas tendrá cuatro horas semanales, entre teóricas y prácticas, según la naturaleza de la enseñanza.	
Al final del tercer año los alumnos presentarán una Memoria para la obtención del título de Diplomado.	

biblioteconómicos que se impartían en la Escuela de Documentalistas, dependiente de la Biblioteca Nacional. A grandes rasgos se podría decir que, salvo excepciones, no existió referencia a los tesauros en nuestro país, hasta finales de la década de los 70. Una de esas excepciones la constituye el seminario que impartió Emilia Currás, en 1969, sobre *La Documentación al servicio de la industria química, farmacéutica e investigación científica*, cuyo programa incluía un epígrafe dedicado a los *thesaurus* (figuras 23 y 24). Tenemos constancia, igualmente, de la celebración de otros cursos y seminarios sobre Documentación científica que, desde el año 1973, y promovidos también por la profesora Currás, se celebraron en el seno de la Facultad de Ciencias, de la Universidad Autónoma. Hacemos mención a continuación de algunos de ellos:

- *Seminario sobre Documentación científica* que, bajo el título *Problemas actuales de la Documentación científica: Caracterización y clasificación de documentos*, se impartió en mayo de 1973 (fig. 25 y 26).

- *Cursillo sobre Documentación química*, celebrado entre los días 10 y 15 de diciembre de 1975 (fig. 27 y 28).

También en el marco de la formación no reglada, José María Sánchez Nistal, vicedirector del CINDOC en la actualidad, impartió el primer curso monográfico sobre tesauros de que tenemos constancia. Se celebró en 1984, y estaba organizado por la Sociedad para el Tratamiento de la Información (SOTRAI), primera empresa privada de documentación en España, que fue creada en 1981.



Instituto  
Nacional  
de  
Racionalización  
del Trabajo

Serrano, 150  
**MADRID**

Figura 23. Programa del Seminario sobre la Documentación al servicio de la industria química, farmacéutica e investigación científica

## PROGRAMA DEL SEMINARIO

1. INTRODUCCION AL TEMA.
  - 1.1. Documentación e Información.
  - 1.2. Documentación y Bibliotecología.
2. CENTRO DE DOCUMENTACION QUIMICA. SU ORGANIZACION.
  - 2.1. Secciones de que debe constar.
  - 2.2. Tipos de catálogos, fichas, formularios, etc., a utilizar.
  - 2.3. Recorrido a seguir por el documento.
3. INFORMACION QUIMICA
  - 3.1. Tipos de información.
  - 3.2. El usuario.
  - 3.3. Formas de hacer llegar la información al usuario.
  - 3.4. Tipos de publicaciones de información química.
4. EL DOCUMENTALISTA.
  - 4.1. Su formación, escuelas, etc.
5. ASPECTOS PRACTICOS DE LA DOCUMENTACION QUIMICA.
  - 5.1. Resúmenes de la Documentación Química.
  - 5.2. Traducciones y su automatización (automatización).
  - 5.3. Sistemas de clasificación.
6. THESAURUS.
  - 6.1. Palabras claves.
  - 6.2. Descriptores.
  - 6.3. Criterio para selecciones, palabras claves.

7. AUTOMACION DE LA DOCUMENTACION QUIMICA.
  - 7.1. Formas de transcribir las fórmulas químicas.
  - 7.2. Confección de catálogos e índices.
  - 7.3. Otras aplicaciones de las calculadoras electrónicas.
8. LA DOCUMENTACION APLICADA A LAS PATENTES.
9. SOCIEDADES NACIONALES E INTERNACIONALES DE DOCUMENTACION.

## DESARROLLO DEL SEMINARIO

El desarrollo de estos temas estará a cargo de doña Emilia Currá, doctora en Ciencias Químicas por las Universidades de Berlín y Madrid. Documentalista por la Escuela Especial de Frankfurt/Main.

El Seminario de unas quince horas de duración se celebrará en Madrid, en los locales del Instituto Nacional de Racionalización del Trabajo, Serrano, 150, de dieciocho a veintiuna horas, ambas inclusive, del 10 al 14 de febrero de 1969.

## ASISTENCIA AL SEMINARIO

Los interesados en tomar parte en el seminario deberán comunicarlo al Departamento de Formación y Asesoramiento del Instituto Nacional de Racionalización del Trabajo, Serrano, 150. Madrid-6. Teléfono 261 70 08, indicando brevemente su posición y cometido dentro de la empresa a que pertenecen.

Una vez les sea comunicada la formalización de su inscripción deberán hacer efectivo en la Sección de Contabilidad del Instituto el importe de los derechos correspondientes, 4.500 pesetas. Los miembros del Instituto sólo abonarán 4.000 pesetas.

Las empresas miembros del Instituto podrán acogerse al mismo beneficio que los miembros individuales, es decir, satisfaciendo 4.000 pesetas por asistente, y con un máximo de tres.

Figura 24. Programa del Seminario sobre la Documentación al servicio de la industria química, farmacéutica e investigación científica

(X) 5.3 Terminología  
5.4 Normalización  
5.5 Abreviaturas



UNIVERSIDAD AUTONOMA DE MADRID  
FACULTAD DE CIENCIAS

SEMINARIO SOBRE DOCUMENTACION CIENTIFICA.

Dr. Julio Garrido  
Dra. Emilia Currás

Día 10 de Mayo a las 16h.

PROBLEMAS ACTUALES DE LA DOCUMENTACION CIENTIFICA  
Caracterización y Clasificación de Documentos

Día 11 de Mayo a las 16h.

BUSQUEDA DE DOCUMENTOS  
Métodos clásicos y métodos mecánicos.  
Documentación e informática.

Departamento de Geología  
Módulo VI, Aula 405.

Madrid, 11 de Abril de 1973.





SEMINARIO SOBRE DOCUMENTACION CIENTIFICA

*Mayo 1973*

BUSQUEDA DE DOCUMENTOS

Métodos utilizados por los investigadores.  
Búsqueda en un conjunto registrado.

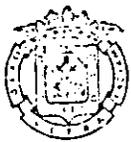
Distribuciones lineales.  
Intersección de distribuciones lineales.

Análisis temático. Concordancias.  
Caracterización de los documentos.  
Tipos de selección.

Método de rejillas superpuestas.  
Métodos semimecánicos.  
Uso de los ordenadores.  
El problema de ideas afines.  
Limitaciones de los ordenadores.

UTILIZACION DE ORDENADORES EN QUIMICA

- 1 .- Aplicación de los ordenadores en Documentación Química.
- 2 .- Maneras de llevar las fórmulas químicas al ordenador.
  - 2.1. Métodos directos.
    - 2.1.1. Uso de la Televisión.
    - 2.1.2. Método topológico.
  - 2.2. Métodos indirectos.
    - 2.2.1. Criterio general.
    - 2.2.2. Método Greimas.
    - 2.2.3. Método Wiswesser.
    - 2.2.4. Otros métodos.
- 3 .- Aplicación de estos criterios a otras ramas de la Ciencia.



UNIVERSIDAD AUTÓNOMA DE MADRID  
FACULTAD DE CIENCIAS

## PROGRAMA DE CURSILLO DE DOCUMENTACION QUIMICA

para los alumnos de Química Analítica

Facultad de Ciencias. Universidad Autónoma de Madrid

### INTRODUCCION

Dado el gran interés que la documentación científica tiene en el campo de la Química se ha considerado importante que los alumnos próximos a salir de nuestra Facultad y que cursan la Asignatura de Química Analítica tengan unas nociones fundamentales de la documentación existente en este campo y de como adquirir información.

Para ello se ha confeccionado un programa en el que se enseñan los tipos de documentos existentes, los sistemas de clasificación, la forma de hacer una referencia bibliográfica, un resumen, como adquirir información, etc.

Todo ello les será de gran utilidad cuando estén en sus puestos de trabajo en la Industria, o en la Investigación, para valerse, en parte, por si mismos y actualizar, así, más rapidamente sus métodos de trabajo y conocimientos.

### CONDICIONES GENERALES

Este cursillo se compone de:

4 clase teóricas de 1 hora de duración

Proyección de películas documentales

Visita a un Centro de Documentación y su terminal de información.

Las clases teóricas irán seguidas de coloquio, y se desarrollarán según el programa que figura a continuación.

### PROGRAMA DEL CURSILLO SOBRE DOCUMENTACION QUIMICA

Lección 1. Documentación e Información Química. Transferencia de Información. Los documentos: Sus tipos. Soportes de la Documentación Química: Clasicos y modernos. Métodos clásicos de clasificación. Mecanización de la documentación. Teoría de los Thesaurus. Palabras-claves. Thesaurus químicos mas importantes.



- Lección 2. Catalogación e Indización en Documentación Química. Como se redacta un trabajo científico, una tesina, una tesis. Resúmenes científicos. Tipos de resúmenes. Referencias bibliográficas. Indización por palabras - claves. Índices en Documentación Química mecanizada.
- Lección 3. Métodos de llevar las fórmulas y las reacciones químicas al ordenador: Métodos de fragmentación. Métodos lineales. Métodos topológicos. Métodos mixtos. Sistema GREMAS. Notación lineal de Wiswesser. Método bidimensional utilizado por el Chemical Abstracts Service. Otros métodos menos utilizados.
- Lección 4. Búsqueda de Información Química. Formulación de preguntas. Donde y como se encuentra la Documentación Química. Información bibliográfica retrospectiva. Difusión Selectiva de Información. Centros de Documentación especializados. Bancos de Datos. Terminales on-line.

Madrid Junio 1975

Dra. Emilia Currás  
Dpto. de Documentación Científica  
Facultad de Ciencias

### 3.3.3. Los primeros tesauros españoles

El primer tesoro que apareció en el panorama documental español fue el *Thesaurus Textil Internacional*<sup>34</sup>, publicado en 1971. Se trataba de la traducción del *Thesaurus del Instituto de Tecnología de Massachusetts*, que había salido a la luz en Estados Unidos en 1966. Dicha traducción formaba parte de un programa de desarrollo y perfeccionamiento de la industria textil, promovido por el Grupo Europeo de Intercambio de Experiencias sobre la Dirección de la Investigación Textil (G.E.D.R.T.), del que formaba parte España. El director de la Escuela Técnica Superior de Ingenieros Industriales de Tarrasa, López Amo, tomó la iniciativa de su adaptación al español, que se llevó a cabo durante un período de más de dos años, por el enorme interés que el glosario podría tener para la industria textil en nuestro país.

Dicho tesoro consta de una lista alfabética de términos clave, que van precedidos del código identificador, y de una lista alfabética compuesta por dichos términos provistos de sus relaciones semánticas con otros que componen el tesoro (figuras 29, 30 y 31).

Pero la traducción de este tesoro constituye un caso aislado, ya que las primeras técnicas documentales no se manifestarían hasta casi finales de la década de los 70, pudiéndose considerar a Lasso de la Vega su precursor, al incorporar en el Centro de documentación de Fundesco -que él mismo fundó- un sistema de indización que bautizó como FUNLA (derivado de Fundesco y Lasso), consistente en fichas similares a las empleadas por el sistema uniterm. Mediante este sistema, se conseguía una recuperación de la información bastante más ágil que con los sistemas tradicionales hasta entonces, basados en los puntos de acceso de autor y título exclusivamente. Pedro Arévalo, que es

---

<sup>34</sup>THESAURUS Textil Internacional. Tarrasa: Escuela Técnica Superior de Ingenieros Industriales, 1971.

# THESAURUS TEXTIL INTERNACIONAL

Edición Española

ESCUELA TECNICA SUPERIOR DE INGENIEROS INDUSTRIALES  
DEPARTAMENTO DE DOCUMENTACION  
TARRASA

TOMO I

A - H

74360	ADARQUILLADO	32020	ACABADO /ASPECTO DE/
33050	ABATANADO	4734	ACABADO /MAQUINARIA DE/
49020	ABERTURA	4737	ACABADO /MACUINA DE/
5530	AULANDAMIENTO	4780	ACABADO /PROCEDIMIENTO DE/
70	ABRASIMETRO		ACABADO /PROCESOS DE/
80	ABRASION	4670	ACABADO A LA CONTINUA
72830	ABRASION /APRESTOS ANTI/	4700	ACABADO BAULK
100	ABRASION /ENSAYO A LA/	65060	ACABADO DE SUPERFICIE /AGENTE DE/
120	ABRASION /RESISTENCIA A LA/	4735	ACABADO EN FASE DE VAPOR
72860	ABRASION /SENALES DE/	4690	ACABADO EN HUMEDO
60	ABRASIVO	4680	ACABADO EN SECO
52275	AURAZADERA		ACABADO MECANICO
49630	ADRE LENGUETAS (DISPOSITIVO)	4720	ACABADO MELTON
5840	AURIDOR	24230	ACABADO PERMANENTE
	ABRIDOR /ORGANO/	4778	ACABADO PLURIVALENTE
49640	ABRIDORA	4720	ACABADO POR EL HAZ
8980	ABRIDORA DE BALAS	48870	ACABADO REPELENTE AL ACEITE
32355	ABRIDORA DE BALAS POR ARRANQUE	32060	ACABADORA /CARDA/
43790	ADRIGO (PRENDA)	4737	ACABADOS /MAQUINARIA PARA/ (EN GENERAL)
42925	ADRILLANTADO	355	ACCESIBILIDAD (INTERNA)
42930	ADRILLANTADORA /MAQ./	360	ACCESORIO
35110	ADRILLANTAMIENTO	33400	ACCESORIOS (CONFECION)
161	ABSORBATO	370	ACCIDENTE
160	ABSORBENTE	14960	ACCIONADORA /CINTA/ (MOVIMIENTO HUSOS)
63490	ABSORBENTE /AGENTE/	42720	ACEITE
220	ABSORBENTE /MATERIA/	36697	ACEITE /BANO DE/
	ABSORCION	36730	ACEITE /MANCHA DE/
165	ABSORCION (ACCION)	36695	ACEITE DE ALIMENTACION
240	ABSORCION (RADIACION)	36700	ACEITE MINERAL
190	ABSORCION /AP. ENSAYO/	36710	ACEITE SULFATADO
63500	ABSORCION /CAPACIDAD DE/	36720	ACEITE SULFONADO
63540	ABSORCION /ENTALPIA DE/	330	ACELERACION (MECANICA)
215	ABSORCION /EQUILIBRIO DE/	340	ACELERACION /PROCESO DE/
250	ABSORCION /ESPECTRO DE/	73240	ACELERADOR (TINTURA)
230	ABSORCION /PODER DE/	320	ACELERANTE (TENIDO)
63560	ABSORCION /VALOR DE/	350	ACELEROMETRO
300	ABSORCION /VELOCIDAD DE/	352	ACCEPTOR
63510	ABSORCION DE AGUA	885	ACERO
63520	ABSORCION DE GASES	38460	ACERO INOXIDABLE
63530	ABSORCION DE LIQUIDOS	500	ACETAL
190	ABSORCION DINAMICA	510	ACETAL /RESINA DE/
185	ABSORCION DINAMICA (AGUA)	520	ACETAMIDA
170	ABSORCION DINAMICA /ENSAYO DE/	530	ACETATO
210	ABSORCION DINAMICA /ENSAYOS DE/	540	ACETATO /COLORANTE AL/
260	ABSORCION ESTATICA	550	ACETATO /ESTER/
270	ABSORCION ESTATICA /CONTROL DE/	560	ACETATO /FIBRA DE/
	ABSORCION ESTATICA /ENSAYO DE/	530	ACETATO /RAYON AL/
310	ABSORTIVIDAD /RADIACION/	580	ACETATO /TENIDO CON CCLORANTE DE/
4800	ACABADO (PROPIEDADES)	62510	ACETATO SODICO
4790	ACABADO /AGENTE DE/	610	ACETICA /SAL/

74360 ABARQUILLADO

03/11/70

PAGINA 1

120 ABRASION /RESISTENCIA A LA/

74360 ABARQUILLADO

POR DESTORSION /TENDENCIA A LA/  
 TENDENCIA A LA DESTORSION  
 TA TEJIDO /DEFECTO DEL/  
 TR TEJIDO DE PUNTO LISO  
 TEJIDO PUNTO CIRCULAR /MAQ. PARA/  
 TEJIDOS DE PUNTO  
 TORSION /ESTABILIDAD DE LA/

33050 ABATANADO

VER BATANADO

49620 ABERTURA

VER APERTURA

5530 ABLANDAMIENTO

TA ACABADO EN HUMEDO  
 TR CALANDRA PARA LAMINADO POR CAPAS  
 LUBRICACION  
 SUAVIZANTE

70 ABRASIMETRO

POR USOMETRO  
 TA ENSAYOS /EQUIPO PARA/

80 ABRASION

POR DESGASTE AL USO  
 TR ABRASION /ENSAYO A LA/  
 ABRASION /RESISTENCIA A LA/  
 ABRASIVO  
 AFILAR (ACCION)  
 ARANAZO  
 BCLITAS  
 COLA (ENCOLADO)  
 CORINDON  
 CORTAR (ACCION)  
 COSTURA /RESIST. A LA ABRASION DE LA/  
 DESGASTE /ACABADO RESISTENTE AL/  
 DESLIZANTE  
 DURABILIDAD  
 ENGANCHADO /RESISTENCIA AL/  
 ENGANCHON /FORMACION DEL/  
 ESCARCHA /FORMACION DE/  
 ESFUERZOS /CONCENTRACION DE/  
 ESMERIL  
 FALDACION /RECTURA POR/  
 FATIGA  
 FATIGA /RESISTENCIA A LA/  
 FRICCION  
 FRICCION /COEFICIENTE DE/

FRICCION DINAMICA  
 GASTO  
 GASTO /RESISTENCIA AL/  
 HILC /DESGLIZAMIENTO DEL/  
 HILC /ROZAMIENTO HILC-CONTRA/  
 HILC /ROZAMIENTO METAL-CONTRA/  
 IRRITACION DE LA PIEL  
 LIMPIEZA (ACCION)  
 LIMPIEZA FUERTE  
 MECANICA /PROPIEDAD/  
 PATINAZO  
 PULIDO  
 RAYADO  
 RAYADO POR FROTACION  
 RAYADO-ARANAZO  
 RAYADURA  
 ROCE  
 RECTURA  
 ROZAMIENTO /FUERZA DE/  
 TEJIDO CONTRA TEJIDO /ROCE/  
 TISAJE /DINAMICA DEL/  
 URDIMERE /TENSION DE LA/

72830 ABRASION /APRESTOS ANTI/

VER DESGASTE /ACABADO RESISTENTE AL/

100 ABRASION /ENSAYO A LA/

POR DESGASTE /ENSAYO AL/  
 TA CONTROL (ENSAYO)  
 TE SCHIEFER /ENSAYO DE ABRASION/  
 STOLL Q.V. /APARATO CONTROL ABRASION/  
 TABER /ENSAYO DE ABRASION/  
 TR ABRASION  
 CCRTE /ENSAYO DE/  
 DESGARRO /ENSAYO AL/  
 DESGASTE /ENSAYO DE RESISTENCIA AL/  
 ECLATOMETRO /ENSAYO AL/  
 ENGANCHON /ENSAYO AL/  
 ENSAYOS /EQUIPO PARA/  
 FATIGA /ENSAYO DE LA/  
 FLEXION /ENSAYO A LA/  
 FLUENCIA /CONTROL DE/  
 FORMACION DE BCLITAS /ENSAYO DE/  
 IMPACTO /ENSAYO DE PENETRACION AL/  
 IMPACTO /ENSAYO DE RESISTENCIA AL/  
 RESISTENCIA /ENSAYO DE LA/  
 TRACCION /ENSAYO DE/

120 ABRASION /RESISTENCIA A LA/

POR DESGASTE AL USO /RESISTENCIA AL/

(CONTINUA)

documentalista en la actualidad en el Centro de Documentación de Fundesco, y *aprendiz* de Lasso de la Vega por aquellos años, según él mismo expresa, nos habla de la creación de una lista de palabras clave en el año 1974, que podría considerarse como el primer indicio de elaboración de un tesaurus en España<sup>35</sup>. Los problemas que se tuvieron que enfrentar para su elaboración, derivan de la dificultad de traducir o adaptar al español conceptos especializados en el área de las comunicaciones, en aquella época; conceptos como *teleinformatique*, por ejemplo, que traducían por *teleinformática* -ahora telemática-, o *software*, que se denominó *soporte lógico*, etc. Resultado de aquella lista de palabras clave fue el *Tesaurus de Fundesco*, que no se publicó hasta 1987, dos años después de que se automatizara la base de datos.

El *Tesaurus de Ciencias de la Información* fue el primero elaborado en nuestro país; lo realizó Antonio García Gutiérrez y figura en el segundo volumen de su tesis doctoral: *Normalización de la documentación informativa: Propuesta de tesaurus español de las Ciencias de la Información*, presentada en 1982, y dirigida por el profesor López Yepes. Se trata, por tanto, de un tesaurus realizado en el marco de una investigación sobre lenguajes documentales, que tras previa actualización editó, en 1985, en tirada limitada, IBERCOM (Centro de Documentación del Departamento de Documentación de la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid). Tuvo un tiempo de aplicación en la Base de datos de Comunicación social de IBERCOM, centro adscrito a la Red COMNET de UNESCO, y posteriormente fue uno de los textos fuente para la construcción del *Tesaurus de Comunicación e Información*, preparado por Jean Viet, auspiciado por UNESCO y la FID.

---

<sup>35</sup> Aunque ésta fue la primera manifestación práctica de creación de un tesaurus, Lasso de la Vega ya se había ocupado de ellos a nivel teórico, con anterioridad. Concretamente en el artículo titulado *Organización de la documentación por palabras y descriptores específicos de materia: los "thesaurus"*, publicado en 1966, en el número 6 de la revista **Racionalización**. En dicho trabajo, el autor ofrece información sobre la elaboración de tesaurus para *la organización de los servicios de información y el empleo de los ordenadores electrónicos*.

### 3.3.3.1. Otras referencias y proyectos de tesauros

Para conseguir datos concretos sobre la construcción de otros tesauros en España recurrimos a fuentes impresas que mencionaremos a continuación, así como a fuentes informales.

De los datos aportados por las fuentes escritas tendríamos que destacar su escasa fiabilidad, en líneas generales, de acuerdo con la falta de rigor de la información. La excepción a esta tónica la constituye un artículo que bajo el título **Desarrollo de lenguajes documentales formalizados en lengua española**, publicó la Revista Española de Documentación Científica en 1989<sup>36</sup>. En él se hace referencia a la existencia de 12 tesauros españoles en esa fecha, cuya existencia pudimos comprobar documentalmente.

Analizamos las ediciones de 1987 y 1994 del **Catálogo de recursos terminológicos detectados en lengua española**, publicado por el CSIC. Este repertorio no nos sirvió de gran ayuda, dado que es quizá la fuente con mayores imprecisiones que consultamos; creemos que los errores que incluye pueden deberse a que su confección se basa en informaciones recibidas en el CSIC mediante encuestas a bibliotecas y centros de documentación, cuyos datos no fueron contrastados, y al uso generalizado del término tesoro para denominar cualquier tipo de vocabulario. Si bien es cierto, peca el mencionado repertorio de otros fallos tales como la repetición de las mismas obras en numerosas ocasiones, y no aporta datos a los que, por otra parte remite, especialmente en lo relativo a los organismos en los que señala la existencia de determinados documentos.

---

<sup>36</sup>ALVARO BERMEJO, C.; VILLAGRÁ RUBIO, A.; SORLI ROJO, A. Desarrollo de lenguajes documentales formalizados en lengua española. II. Evaluación de los tesauros disponibles en lengua española. *Revista Española de Documentación Científica*, v. 12, nº 3 (1989) pp. 283-301.

Dadas estas omisiones, no fue tarea fácil dar con las fuentes originales que aportaran la información exacta y documentada de los supuestos tesauros. Queríamos comprobar uno a uno todos los datos que nos ofrecía el Catálogo. De los mencionados *tesauros* por dicho repertorio, no todos resultaron serlo. A continuación mostramos algunos ejemplos de ello:

- En relación con un denominado *Tesaurus de Ciencias sociales*, Silvia Serrano, documentalista del Centro de Información y Documentación de Barcelona (CIDOB), a donde remite el Catálogo de Recursos Terminológicos para su localización, nos confirmó su inexistencia. En dicho Centro indizan únicamente con los siguientes tesauros: el de Unesco, el de la OIT y los de Urbanismo, Economía y Topónimos del ISOC.

- En cuanto a un *Tesaurus de Comercio Interior*, publicado con fecha 1972, detectamos la imprecisión en el momento en que vimos dicha fecha de publicación, a todas luces demasiado temprana en lo que a edición de tesauros en España se refiere. Consultamos en la biblioteca dependiente de la Dirección General de Comercio Interior, al haber dejado de existir en los años 80 el Instituto de Reforma de la Estructuras Comerciales (IRESCO) -institución a la que remite el Catálogo de recursos terminológicos-. En la biblioteca de Comercio Interior nos aseguraron que nunca existió tal tesaurus, y que debía tratarse de un vocabulario que se elaboró en un centro de documentación sobre comercio interior que se creó en el IRESCO, y que estaba basado en los términos que recogían de la información que iba llegando al Instituto.

- En el Instituto Eduardo Torroja de la Construcción y el Cemento, nos confirmaron que nunca ha existido un *Thesaurus de la Construcción y Obras Públicas*. Nos inclinó una vez más a efectuar la comprobación la fecha -1972- que recoge el Catálogo de recursos terminológicos. La única herramienta de indización con que han

contado en el mencionado Instituto es un *Léxico de la construcción*<sup>37</sup>, con depósito legal fechado en 1968.

- Nos referimos, por último, al *Thesaurus español-francés de Ciencias de la Tierra*, del Instituto Geológico y Minero de España, en la actualidad denominado Instituto Tecnológico Geominero de España. Alonso Sánchez, documentalista del centro, nos comunicó la existencia de un amplísimo léxico de más de 15.000 términos, de actualización periódica, cuya última puesta al día se realizó este mismo año.

Otra fuente consultada fue el **Centro Argentino de Información Científica y Tecnológica (CAICYT)**, organismo que centraliza la información relativa a publicación de tesauros en lengua española. El Centro tan sólo tiene registradas dos versiones en castellano de tesauros internacionales, concretamente el *Tesauro de carreteras*, del Ministerio de Obras Públicas y el *Spines*, de la Unesco; y cinco tesauros españoles: el *Tesauro de Servicios sociales*, del Ministerio de Trabajo y Seguridad social, en su edición de 1986, el *Tesauro ISOC de Urbanismo*, en edición de 1988, *Tesauro ISOC de topónimos* (1991), *Tesauro ISOC de Psicología* (1992) y *Tesauro ISOC de Economía* (1992)<sup>38</sup>. Como puede observarse, los lenguajes que figuran en el CAICYT tienen fecha reciente, lo cual no aporta información de utilidad sobre el tema de estudio.

La consulta a la **Base de datos Thesauri**, del distribuidor de la Comunidad Europea Echo nos dió un resultado negativo, ya que no tienen registrada en dicha base la existencia de ningún tesauro español.

---

<sup>37</sup>INSTITUTO EDUARDO TORROJA DE LA CONSTRUCCIÓN Y EL CEMENTO. *Léxico de la construcción*. Madrid: I. E.T.C.y C., 1968.

<sup>38</sup>La referencia bibliográfica completa de estos tesauros se incluye en el epígrafe 3.4., comprendido en este mismo capítulo.

En cuanto a **catálogos de bibliotecas**, obtuvimos informaciones repetidas, confusas y poco fiables, en términos generales. Una vez más tenemos la sospecha de que la imprecisión terminológica con que se emplea el término tesoro es la causa de dichas incorrecciones.

Del análisis de los **repertorios y directorios** a los que tuvimos acceso, apenas se obtuvieron referencias de tesauros construidos en España. Algunos de estos repertorios se referían a tesauros de difusión internacional, con edición española, pero no realmente a tesauros elaborados en nuestro país. Los directorios consultados fueron: **Bibliographie des thesauri et index par matières parus depuis 1960**<sup>39</sup>; **Thesaurus guide: Analytical directory of selected vocabularies for information retrieval**<sup>40</sup>, publicado por las Comunidades Europeas, y que revisamos desde la edición de 1985; **Bibliografía de vocabularios, tesauros, encabezamientos de materia y esquemas de clasificación de ciencias sociales**<sup>41</sup>; **Thesauri used in online databases: An Analytical Guide**<sup>42</sup>; **Bibliography of United Nations Thesauri, classifications, nomenclatures**<sup>43</sup> y, finalmente, **Annuaire des thésaurus suivi d'un index par domaines et sous-domaines**

---

<sup>39</sup>LAUREILHE, M.T. **Bibliographie des thesauri et index par matières parus depuis 1960**. Paris: Bibliothèque Nationale, 1975.

<sup>40</sup>THESAURUS **guide: Analytical directory of selected vocabularies for information retrieval/prepared by Gesells Chaft für Information und Dokumentation for the Commission of the European Communities**. Amsterdam: North-Holland; Luxembourg: Office for Official Publications of the European Communities, 1985.

<sup>41</sup>BIBLIOGRAFÍA **de vocabularios, tesauros, encabezamientos de materia y esquemas de clasificación de ciencias sociales (mono y plurilingües)**. Preparado por la Biblioteca Aslib en colaboración con Jean Aitchison y C. G. Allen. Paris: Unesco, 1982.

<sup>42</sup>CHAN, L.M. & POLLARD, R. **Thesauri used in online databases: An Analytical Guide**. New York, etc.: Greenwood Press, 1988.

<sup>43</sup>BIBLIOGRAPHY **of United Nations Thesauri, classifications, nomenclatures**. Geneva: Inter-Organization Board for Information Systems, 1979.

couverts par les thésaurus et d'une Liste des Organismes auteurs<sup>44</sup>, preparado en Francia por el Bureau National d'Information Scientifique et Technique.

El único repertorio en el que se registra información relativa a España es en la **Bibliografía internazionale di thesauri**<sup>45</sup>. En ella se hace referencia a un *Thesaurus jurídico inteligente. Proyecto trienal de la Universidad de Zaragoza para la recuperación automática de las normas de la Región de Aragón* y a un *Thesaurus del sistema di documentazione dei trattati internazionali/International treaties retrieval system thesaurus*, denominado *IBERTRAT*, editado por el Ministerio de Asuntos Exteriores, ambos sin fecha de publicación. El primero consiste en una lista de términos recopilados con objeto de crear un sistema experto en materia de Derecho. Al frente del proyecto estaba Emilio Suñé, quien presentó una ponencia en el "4º Congresso Internazionale sul tema: Informatica e Regolamentazioni Giuridiche", celebrado en Roma en 1988. En la ponencia, titulada *Los tesauros en bases de datos de usuario*, se da la siguiente definición de tesoro: ... *es un conjunto de descriptores, que tienen la posibilidad de estar aclarados en su alcance, relacionados de diversas formas entre sí, y aglutinados esencialmente bajo un criterio alfabético de ordenación*. A esta definición le sigue un párrafo, en el que el autor se expresa en los siguientes términos: *Por extensión, emplearemos también el término tesoro para definir cualquier lista de palabras, incluso cuando sean obtenidas automáticamente, que nos den acceso al fondo documental de la base de datos*.

Este uso de la palabra tesoro es ilustrativo de la confusión que ha existido a lo largo de los años, y sigue existiendo en la actualidad, a la hora de diferenciar un tesoro

---

<sup>44</sup>BUREAU NATIONAL D'INFORMATION SCIENTIFIQUE ET TECHNIQUE. *Annuaire des thésaurus suivi d'un index des domaines et sous-domaines couverts par les thésaurus et d'une Liste des Organismes auteurs*. 3ª ed. Paris: Comité de Coordination des Thesaurus Sectoriels, 1975.

<sup>45</sup>RIZZO, S. *Bibliografía internazionale di thesauri/International Bibliography of thesauri*. Roma: Camera dei deputati, 1987.

documental de una lista de palabras clave o de cualquier otro lenguaje de recuperación de información.

En cuanto al denominado *Tesouro IBERTRAT*, consiste en una lista de palabras clave que recoge el nombre de los tratados suscritos por el Ministerio de Asuntos Exteriores con otros países de habla hispana, que en ningún momento se llegó a editar.

Dada la inexistencia de otras fuentes impresas en las que recabar algún punto de referencia fiable sobre los primeros tesauros españoles, intentamos la conexión con fuentes informales.

En conversaciones mantenidas con especialistas del área, obtuvimos las informaciones que se detallan a continuación:

En opinión de Sánchez Nistal, no existieron tesauros de creación española hasta principios de los años 80. Esta carencia se suplía en el ISOC con el uso de tesauros multinacionales para recuperar en bases de datos internacionales. Parece, sin embargo, que la necesidad urgente de manejar útiles propios para la recuperación e intercambio de información impulsó la creación de tesauros propios en materia de ciencias sociales y tecnología. El primer tesoro que se elaboró con este objetivo fue el de urbanismo, que ya ha sido objeto de revisión en fecha posterior. Los primeros lenguajes especializados eran financiados por proyectos de investigación y nacieron paralelamente al desarrollo de las bases de datos.

También en los años 80 se empezó a trabajar en programas de gestión de bases de datos, que incluían, en teoría, gestión de tesauros. Aunque no se trataba de programas específicos para elaborar estos lenguajes, ayudaba a hacerlos y a explotarlos. Hasta 1989

no se dispuso, sin embargo, de un programa específico de elaboración automática de tesauros en España. Antonio Avia fue el primero en hacer una aplicación informática de este tipo. Lo llevó a cabo en el programa de gestión de bases de datos para microordenadores dBASEIII, que está actualmente en vigor; se trata del programa ANDAMIO, que consiguió poner en marcha tras varias pruebas realizadas en 1988. Se ceñía estrictamente a las fases de elaboración que estaban previstas en el ISOC, a esta ventaja se unía la de estar pensado en exclusiva para la lengua española y la de permitir el trabajo simultáneo en varios tesauros al mismo tiempo, sin limitación de memoria (para número de descriptores o número de caracteres por descriptor), a diferencia de otros programas.

Amelia de Irazazábal, componente del grupo de terminología del ICYT, sostiene que en 1970 se empezaba a hablar y a trabajar con estos lenguajes documentales, *cada uno en su despacho desarrollaba sus propias experiencias en materia de elaboración de tesauros, lo mejor que sabía, sin ningún género de coordinación o normalización*. Coincidiendo con aquella fecha el Instituto de Racionalización y Normalización (IRANOR) -que por aquel entonces pertenecía al CSIC-, conjuntamente con el Centro de Cálculo, de la misma institución, hicieron la versión española del *Tesaurus de la Unesco*. También por esos años se hizo la adaptación española del Tesaurus de política científica del mencionado organismo, el *Tesaurus Spines*. Esta fue quizá la tarea más comprometida en aquella época en materia de tesauros, dado que el Spines es un tesaurus de grandes dimensiones, compuesto por un campo terminológico muy amplio. De cada uno de los 33 campos temáticos que lo componen se hizo cargo un especialista, quien tradujo, adaptó y normalizó su terminología de forma independiente, procediéndose después a la unificación de todos los campos.

Amelia de Irazazábal nos facilitó, igualmente, un informe inédito, fechado en el CSIC, en 1982, bajo el título *Evolución de tesauros en España. Algunos problemas en*

*su desarrollo*<sup>46</sup>. Dicho informe hace referencia a la Conferencia Iberoamericana sobre Información y Documentación Científica y Tecnológica, REUNIBER'78, celebrada en Madrid, con la colaboración de la Unesco y de la OEA, y organizada por el ICYT. En la Conferencia se trataron de sentar las bases para la coordinación de esfuerzos en materia de construcción de tesauros, poniéndose de manifiesto *la semejanza de los problemas existentes en el campo de los lenguajes documentarios controlados en lengua española*<sup>47</sup>. En el marco de REUNIBER'78 se creó un grupo de trabajo sobre Lenguajes documentarios coordinado por Argentina, y con secretaría en el Centro Argentino de Información Científica y Tecnológica (CAICYT). *Este grupo publicó en 1981 un inventario de lenguajes documentarios no tradicionales en lengua española y/o portuguesa*,<sup>48</sup> entre los que figuraban once tesauros desarrollados en España, que se mencionan en el informe del CSIC, y cuyas características comentamos a continuación. Antes de analizar dichos datos queremos, no obstante, advertir de la existencia de imprecisiones del concepto de tesoro, patentes en el informe, fruto probablemente de la escasa información y bibliografía en español, en esa fecha, sobre los lenguajes postcoordinados. Es común, por ejemplo, identificar un tesoro con un léxico, que no es sino una simple lista que respeta un orden lexical; también se le asocia con un índice, es decir, con listas organizadas en las que cada concepto envía a las referencias. Por último, se le confunde con un diccionario, con el que tiene diferencias estructurales, de contenido y de finalidad.

---

<sup>46</sup>IRAZAZÁBAL NERPELL, A. *Evolución de tesauros en España. Algunos problemas en su desarrollo*, [inédito]. Madrid, 1982.

<sup>47</sup>IRAZAZÁBAL NERPELL, A. Op. cit.

<sup>48</sup>IRAZAZÁBAL NERPELL, A. Op. cit.

Las referencias del informe son las siguientes:

**1. *Tesaurus sobre aplicación de telecomunicaciones al desarrollo***  
*(en proyecto). Fundesco, a cargo de Lasso de la Vega.*

En la fuente documental no existe fecha de referencia de este proyecto. El documentalista de Fundesco Pedro Arévalo nos facilitó la fecha de 1985 para su realización y 1987 para su publicación, como mencionamos anteriormente.

**2. *Tesaurus sobre Artes del espectáculo. Español y catalán. Barcelona: Biblioteca y Museo del Instituto del Teatro. Irene Peypoch i Mani.***

El denominado *Tesaurus sobre artes del espectáculo* tuvo su origen en un catálogo de materias y conceptos que se empezó a elaborar en 1970. En 1988 dicho catálogo se pasó a formato de papel, presentándose como una lista de descriptores coordinados. Según nos indicaron en la propia biblioteca, se hizo un intento de construcción de un tesaurus en base a esos descriptores, pero sólo llegó a crearse un borrador de carácter experimental. Los responsables de la biblioteca desconocen la existencia de algún ejemplar para poder comprobar con exactitud sus características estructurales. En cualquier caso, el supuesto tesaurus nunca se aplicó al tratamiento de la documentación, ni se publicó.

**3. *Grasas y detergentes. (Lista de términos)*<sup>49</sup>. Instituto de la Grasa. CSIC, Sevilla.**

En 1970, en el Instituto de la Grasa se empezaron a asignar palabras clave (*descriptores*) a cada uno de los documentos -que denominan *documentos originales* - publicados en la revista bimestral *Grasas y aceites*, editada por el mencionado instituto. Dichos *descriptores* se tomaban de un *Tesaurus de Materias grasas y sus derivados*, elaborado por los servicios bibliográficos de la institución. Analizadas las fuentes (figura 32) comprobamos que el denominado *Tesaurus de Materias grasas y derivados* consiste en una lista de palabras clave sin relaciones semánticas entre ellas y sin organizarse en ningún tipo de estructura en particular. La ausencia de una estructura tesaurizada y su disposición en un fichero manual permite la actualización de la lista, de publicación anual, que se va nutriendo de nuevos términos a medida que van apareciendo en el *Índice de "palabras clave" de los "documentos originales"*, pero que, en ningún caso, constituye un tesaurus.

**4. *Lista de descriptores de química industrial, electrotecnia y metalurgia. ICYT.***

Ninguna persona relacionada con el tema, tanto del ICYT, como de empresas eléctricas, tiene conocimiento de su existencia.

**5. *OCDE - DIC (1974)***

Cuando iniciamos la recopilación de datos acerca de este tesaurus partíamos de la sospecha de que la fecha señalada en el informe no podía corresponder a una versión

---

<sup>49</sup>La información relativa a esta referencia la conseguimos directamente de Carlos Gómez Herrera, promotor de la lista, actualmente jubilado, quien gentilmente nos explicó los pormenores sobre el origen del *Tesaurus de materias grasas y sus derivados*.

<u>Fig.</u>	<u>Palabras-clave</u>	<u>Restantes palabras-clave del mismo "documento original"</u>
74	Aceite de hígado de bacalao	Acido graso poliinsaturado (obtención) – Método de los compuestos de inclusión de urea.
08	Aceite de huicungo (caracterización)	<i>Astrocaryum macrocalyx</i> – Esterol – Glicérido.
49	Aceite de oliva	Aceite de girasol – Aceite de girasol alto oleico – Antioxidante – Enranciamiento.
79	Aceite de oliva	Análisis estadístico – Estudio colaborativo – Parámetro de pureza.
55	Aceite de oliva (extracción)	Decantador centrífugo horizontal – Transformación (rendimiento)
99	Aceite de oliva virgen (extracción)	Parámetro de calidad – Sistema continuo en dos fases – Variables del proceso (influencia de).
00	Aceite de salvado de arroz	Arroz precocido – Secado al aire (efecto del) – Secado en microondas (efecto del).
96	Aceite de semilla	Acido graso (composición) – <i>Brunfelsia uniflora</i> (Solanaceae).
1	Aceite de semilla	Desierto de Sonora – Estudio físico-químico – Leguminosa – Méjico
33	Aceite de semilla de calabaza	Calentamiento continuo – Calentamiento intermitente – Propiedades físico-químicas (cambios).
99	Aceite de semilla de uva	Acido graso – Dióxido de carbono supercrítico – Extracción supercrítica.
35	Aceite de soja	Aceite esencial – Efecto antioxidante.
36	Aceite de soja	Calidad – Información tecnológica – Producción.
37	Aceite de soja	Formulación de productos – Información tecnológica.
35	Aceite esencial	Aceite de soja – Efecto antioxidante.
36	Aceite vegetal	Cromatografía gaseosa – Escualeno (determinación).
38	Aceite vegetal	Emulsión – Ester de sacarosa – Mayonesa – Viscoelasticidad – Viscosidad.
34	Aceite vegetal	Grasa vegetal – Propiedades físicas.
31	Aceituna verde de mesa	<i>Ascolana tenera</i> var. – Cromatografía de afinidad sobre cobre inmovilizado – Fenolasa (fraccionamiento).
33	Aceituna verde de mesa	Color – Conservación (efecto de) – Chalkidiki (variedad) – Tratamiento habitual (efecto del).
34	Acido fenólico	Semilla (humedad) – Torta de semilla de colza.
5	Acido fosfatídico	Actividad tirosina quinasa – Diacilglicerol quinasa – Fosfolipasa D – Información (artículo) – Receptor.
9	Acido graso	Aceite de semilla de uva – Dióxido de carbono supercrítico – Extracción supercrítica.
3	Acido graso (composición)	Aceite de semilla – <i>Brunfelsia uniflora</i> (Solanaceae).
2	Acido graso (contenido en)	Aceite de aguacate – Aguacate – Maduración (influencia de).
1	Acido graso libre	Cromatografía en capa fina/detector de ionización de llama – Destilado de desodorización – Esterol – Hidrocarburo – Triglicérido.
4	Acido graso poliinsaturado (obtención)	Aceite de hígado de bacalao – Método de los compuestos de inclusión de urea.
4	Acidos grasos esterificados (composición)	Cuajo animal – Cuajo vegetal – Queso de Los Pedroches.
0	Actividad antimicrobiana	Agente tensioactivo anfotérico – Estructura química – Propiedades tensioactivas.
5	Actividad tirosina quinasa	Acido fosfatídico – Diacilglicerol quinasa – Fosfolipasa D – Información (artículo) – Receptor.
0	Agente tensioactivo anfotérico	Actividad antimicrobiana – Estructura química – Propiedades tensioactivas.

Figura 32. Tesauro de Materias grasas y derivados

española, por parecernos prematura. La versión más antigua del *Tesaurus DIC* en lengua española parece ser la editada en 1982, que se corresponde con la edición internacional de 1979 más las puestas al día de 1981 y 1982.

El *Tesaurus DIC* tiene su origen en la base de datos DIC (Documentación Internacional de Carreteras) y su principal característica externa es la forma de presentación gráfica flechada. La versión de 1982 corrió a cargo de FUINCA (Fundación de la Red de Información Científica Automatizada) y fue realizada por el grupo de expertos formado por Juan Ignacio Sacristán de Lama, M<sup>a</sup> Teresa Iglesias del Pozo y Juan Ignacio Cuesta Benito.

**6. *Tesaurus sobre medicina del trabajo, seguridad e higiene (en elaboración). Servicio social de higiene y seguridad en el trabajo. Instituto Territorial de Barcelona.***

Consultadas las fuentes oportunas, averiguamos que no existe ningún tesaurus sobre la materia. La información responde al error tipográfico de una revista que editó el Instituto Territorial de Barcelona, en la que se anunciaba un curso sobre elaboración de tesaurus. Se trataba, en realidad, del *Tesaurus de la Organización Internacional del Trabajo (OIT)*. En las figuras 33 y 34 se pueden observar la errata y su enmienda en edición posterior.

**7. *Nomenclatura codificada de diagnósticos (base para un tesaurus de terminología clínica). Facultad de Medicina CSIC, Centro de Documentación de Informática Médica. Valencia (1972).***

Se trata de un amplio listado de términos sin ningún tipo de relación semántica entre ellos, organizados por número currens, con cinco dígitos, que hacen posible la

ampliación de la terminología. La obra presenta dos partes: la primera, numérica, mantiene las grandes categorías de la *Clasificación Internacional de Enfermedades* de la Organización Mundial de la Salud, y reúne más de 14.000 formas diagnósticas. La segunda consiste en un índice alfabético (figuras 35 y 36).

Su autor, Eduardo Agustí, tenía la intención -según informaciones de Julia Osca, responsable de la biblioteca de la Facultad de Medicina del CSIC- de elaborar un tesoro que no llegó a realizar, y nadie retomó su iniciativa.

El compendio constituye una riquísima base terminológica que, por sí sola, facilita el almacenamiento y recuperación de información, pero que podría transformarse en una herramienta más eficaz, previo establecimiento de relaciones semánticas entre los términos que recoge en sus 697 páginas.

**8. *Versión española del Thesaurus de l'Association Française pour l'étude des eaux.*** (en elaboración). Tarrasa. Asociación nacional de ingenieros industriales. Departamento de Documentación.

El único lugar en el que conocen algún dato sobre la existencia del thesaurus es en el Colegio de Ingenieros de Cataluña, en donde tenían una vaga idea de que una persona, jubilada en la actualidad, se encargó de realizarlo, pero no supieron darnos más información al respecto, ni localizaron ningún ejemplar.

Los datos obtenidos de personas relacionadas con el tema permiten confirmar que la versión española del tesoro no llegó a realizarse, quedando en simple proyecto.



1ª Edición

- 6.1.7.- SOERGEL, D.  
Indexing Languages and Thesauri: Construction and Maintenance  
Los Angeles, Melville Publishing Co., 1974
- 6.1.8.- UNESCO  
Principles directeurs pour l'établissement et le développement de thesaurus scientifiques et techniques monolingues destinés à la recherche documentaire  
Paris, UNESCO, 1971
- 6.2.- Tesauros depositados en la biblioteca del C.N.I.D.
- 6.2.1.- Thesaurus of structured terms in occupational safety and health  
Rockville, NIOSH, 1971
- 6.2.2.- UNESCO Thesaurus  
Comilat per AITCHISON, J.  
Old working, Union Brothers, 1977, 2 vols., ISBN-42-3-101469-2
- 6.2.3.- O.I.T.  
Tesauro O.I.T. ( Tesauro de los descriptores empleados para el tratamiento de la información ) en la biblioteca del C.N.I.D. → error corrigido en la 1ª ed de la edición OIT  
Ginebra, O.I.T., 1976, ISBN 92-2-001264-2
- 6.2.4.- CENTRE INTERNATIONAL D'INFORMATIONS DE SECURITE ET D'HYGIENE DU TRAVAIL (CIS)  
Thesaurus CIS. Sécurité et Hygiène de Travail  
Geneve, CIS, O.I.T., 1976, ISBN 92-9096-201-5
- 6.2.5.- INTERNATIONAL ATOMIC ENERGY AGENCY  
INIS: Thesaurus. IAEA - INIS - 13 ( Rev. 6 )  
Vienna, I.A.E.A., 1974

→ error corrigido en la 1ª ed de la edición OIT

NORA

Cent

Bibliografía índices e indización

VICKERY, B. C.

Techniques of information retrieval

London, Butterworth, 1970

CUADRA, A.

Annual Review of Information Science and Technology

New York, John Wiley, Vol. 2, 1967, Vol. 3, 1968

Chicago, Am. Soc. for Inform. Science, Encyclopedia Britanica, Inc., Vol. 5, 1970

Washington, Am. Soc. for Inform. Sci., vol. 7, 1972

COLLISON, R. L.

Indexes and indexing (4ª ed.)

London, E. Benn. Ltd., 1972

Bibliografía tesauros

VICKERY, B. C.

Techniques of information retrieval

London, Butterworth, 1970

AITCHISON, J., A. GILCHRIST

Thesaurus construction - A practical manual

London, ASLIB, 1972

ISO

Documentación. Principios básicos para la redacción y desarrollo de tesauros monolingües

ISO 2788 1974

SOERTEL, P.

Indexing languages and thesauri: Construction and maintenance

Los Angeles, Melville Publishing Co., 1970

CIS

Thesaurus CIS, Sécurité et Hygiène du Travail

Geneve, CIS, 1976, ISBN 92-9096-201-3

AITCHISON, J.

IJNESCO Thesaurus

Old working, Union Brothers, 1977

OIT

Tesauro OIT (Tesauros de los descriptores empleados para el tratamiento de la información en la Biblioteca de la OIT)

Ginebra, OIT, 1976, ISBN 92-2-001264-2

INTERNATIONAL ATOMIC ENERGY AGENCY

INIS: Thesaurus IAEA - INIS-13 (Rev. 6)

Vienna, I.A.E.A., 1974

Figura 34. Extracto de una publicación del Instituto Nacional de Seguridad e Higiene en el Trabajo

EDUARDO AGUSTÍ

*Profesor Agregado interino de Documentación Médica y Subdirector del  
Centro de Documentación e Informática Médica.  
Facultad de Medicina de Valencia*

NOMENCLATURA CODIFICADA  
DE DIAGNÓSTICOS (N. C. D.)  
BASES PARA  
UN *THESAURUS* DE TERMINOLOGÍA CLÍNICA



CENTRO DE DOCUMENTACIÓN E INFORMÁTICA MÉDICA  
FACULTAD DE MEDICINA  
PATROCINADO POR LA CAJA DE AHORROS Y M. DE P. DE VALENCIA  
VALENCIA, 1972

57700	ENFERMEDADES DEL PANCREAS MALADIES DU PANCREAS DISEASES OF PANCREAS KRANKHEITEN DER BAUCHSPEICHELDRÜSE
57711	NECROSIS INFECCIOSA DEL PANCREAS
57721	HEMORRAGIA PANCREATICA
57731	PANCREATITIS
57732	PANCREATITIS AGUDA
57733	PANCREATITIS CRONICA
57734	PANCREATITIS PURULENTA
57735	ABSCESO DEL PANCREAS
57736	PANCREATITIS HEMORRAGICA
57737	PANCREATITIS CRONICA - PANCREATITIS INTERSTICIAL
57739	PANCREATITIS CRONICA RECIDIVANTE
57741	QUISTE PANCREATICO
57751	HIPEREMIA PANCREATICA
57761	CALCULO PANCREATICO
57763	CONCREMENTO PANCREATICO
57771	ATROFIA PANCREATICA
57773	DEPOSITO PIGMENTARIO PANCREATICO
57775	FIBROSIS PANCREATICA
57777	CIRROSIS PANCREATICA
57781	NECROSIS PANCREATICA
57782	NECROSIS AGUDA PANCREATICA - NECROSIS PANCREATICA HEMORRAGICA - AUTODIGESTION PANCREATICA - PANCREOPATIA AGUDA DEGENERATIVA - PANCREATITIS AGUDA ASEPTICA - PANCREOPATIA ENZIMATICA AGUDA - PANCREATITIS HEMORRAGICA AGUDA - APOPLEJIA PANCREATICA - PANCREATOSIS AGUDA
57783	NECROSIS PANCREATICA GRASA
57784	NECROSIS PANCREATICA SEPTICA
57791	AQUILIA PANCREATICA
57795	INSUFICIENCIA PANCREATICA
58000	NEFRITIS AGUDA NEPHRITE AIGUE ACUTE NEPHRITIS AKUTE NEPHRITIS
58011	ENFERMEDAD DE BRIGHT AGUDA - NEFRITIS AGUDA
58013	GLOMERULITIS AGUDA
58015	GLOMERULONEFRITIS AGUDA
58021	NEFRITIS CATARRAL AGUDA
58022	NEFRITIS EXUDATIVA AGUDA
58023	GLOMERULONEFRITIS HEMORRAGICA AGUDA - NEFRITIS HEMORRAGICA AGUDA

Figura 36. Nomenclatura codificada de diagnósticos

### **9. Tesauro ISO. IRANOR.**

En el Instituto de Racionalización y Normalización (IRANOR) se empezó a utilizar en 1983 el *Thesaurus International Technique (TIT)*, especialmente diseñado para bases de datos normativas y reglamentarias, editado por la Association Française pour la Normalization (AFNOR). Dicho tesauro, compuesto por un índice alfabético estructurado provisto de relaciones semánticas entre sus términos, se corresponde con una 2ª edición del publicado por AFNOR en 1980, del que no se recibiría copia en España hasta ese año, y que incluía las modificaciones surgidas entre 1980 y 1983.

Se comenzó a traducir en dicha fecha en IRANOR y en el Consejo Superior de Investigaciones Científicas, con la colaboración de AFNOR, y se procedió a su impresión, para uso interno, en el ICYT. Antonio Valle, especialista en el campo de la informática del mencionado Instituto, refiere los años 1977-1978 como la fecha aproximada en la que se comenzó a realizar la versión española informatizada del *Tesauro ISO*. En esos años prepararon y planificaron la mencionada versión con los medios informáticos con que contaban: un ordenador IBM 360/44, siendo la entrada por cinta magnética. Por esa época, únicamente se disponía de una impresora que sólo tenía el tren de mayúsculas. El recurso utilizado para hacer distinción tipográfica era hacer varios golpes de impresión repetidos sobre el carácter o caracteres a destacar, ya que no existía fórmula de impresión en negrita. Como no había letra grande, configuraban un alfabeto en una matriz de 7 por 7 y, por este sistema, dibujaban dicho alfabeto que mediante una rutina, expandía la letra y se imprimía sobre la página. Ese fue el procedimiento empleado para realizar la portada de la versión informatizada del tesauro de referencia (figuras 37 y 38).

T T T T	H H	E E E E	S S S S	A A A	U U	R R R R	U U	S S S S
T	H H	E	S	A A	U U	K R	U U	S
T	H H H H	E E E E	S S S S	A A	U U	K R R R	U U	S S S S
T	H H	E	S	A A A A A	U U	R R	U U	S
T	H H	E E E E	S S S S	A A	U U U	R R	U U U	S S S S

E E E E	S S S S	P P P P	A A A	U U U	U U	L
E	S	P P P	A A	N N N	U U	L
E E E E	S S S S	P P P P	A A	N N N N	U U	L
E	S	P	A A A A A	N N N	U U	L
E E E E E	S S S S S	P	A A	N N	U U	L L L L L

\* \* \*  
 \* \* \*  
 \* \* \* \* \*  
 \* \* \*  
 \* \* \*

Figura 37. Tesouro ISO. IRANOR

## E1-12-PIRIL (AZO)-2-NAFTOL

008 025

- = 0764490
- < REACTIVO QUIMICO

## ARD

002

- = 0666005
- < FECHA DEL CALENDARIO
- FECHA DE FABRICACION
- HORA
- DIA
- SEMANA

## AÑO DE FABRICACION

- > FECHA DE FABRICACION

## ABACA

- = 0065280
- < FIBRA LIBERIANA
- CARAMO
- YUTE
- SISAL

## ABACO

- > DIAGRAMA

## ABASTECIMIENTO

- > APROVISIONAMIENTO

## ABASTECIMIENTO DE AGUA

041

- = 0326155
- ALIMENTACION (DE MAQUINAS)
- CANALIZACION DE AGUA
- DISTRIBUCION DE AGUA
- AGUA DE ALIMENTACION INDUSTRIAL
- AGUA POTABLE
- EVACUACION DE AGUAS
- PRODUCCION DE AGUA

## ABASTECIMIENTO DE AREAS URBANAS

014

- = 1065012
- AREA URBANA

## ABASTECIMIENTO DE GAS

068

- = 0865458
- ALIMENTACION (DE MAQUINAS)
- CUARTO DE CALDERAS
- DISTRIBUCION DE GAS

## ABCISA

- > COORDENADA

## ABDOMEN

027

- = 1145221

## ABERTURA

033 040

- = 0370569
- ACCESO
- VAVO
- ABERTURA DE VENTILACION
- BOCALLAVE
- VENTANA
- GUARNICION PARA CERCOS
- SALICA
- TRAGALUZ
- OFICIO
- ABERTURA: OPTICA
- ABERTURA DE ACCESO
- ABERTURA DE INSPECCION
- PUERTA

## ABERTURA DE ACCESO

034 048 087 089 092

- = 0618220
- CAMION CISTERNA
- ABERTURA
- PUERTA
- DEPOSITO
- VAGON CISTERNA

## ABERTURA DE HOMBRE

- > ABERTURA DE ACCESO

## ABERTURA DE INSPECCION

008 030 101

- = 0619582
- ABERTURA
- DEPOSITO

## ABERTURA DE REVISION

- > ABERTURA DE INSPECCION

## ABERTURA DE VENTILACION

033 034

- = 0789083
- ABERTURA
- CUBIERTA

## ABERTURA RELATIVA

095

- = 0506686
- > ABERTURA: OPTICA

## ABERTURA: OPTICA

095

- = 0045090

A las dificultades técnicas apuntadas se sumaron las derivadas de la traducción de los términos del tesoro; si bien es cierto que hablar de traducción de un tesoro no es hablar en términos precisos, en ningún caso es posible traducir los descriptores o palabras clave que componen un repertorio terminológico, se trata más bien de realizar una equivalencia idiomática que engendra serias dificultades. En el caso del tesoro que nos ocupa, el establecimiento de dicha equivalencia siguió, en líneas generales, los siguientes pasos:

1. Se extrajeron los términos diferentes, solos o con adendas o datos adicionales tales como la temática, por ejemplo, ya que en el idioma original un mismo término formaba parte de diferentes campos semánticos, pudiendo ser en español traducible en forma diferente en cada tema. Por ejemplo, el término inglés *trainer* puede tener como equivalentes en español aprendiz en un oficio, estudiante, etc. Se dió en ocasiones el caso contrario: que diferentes términos en otro idioma tenían equivalencia en una sola palabra española, por ejemplo, *tacto*, que en nuestro idioma puede entenderse como tino, acierto o sentido del tacto.

2. Se distribuyeron los términos diferentes entre expertos para que los circunscribieran y transcribieran a la equivalencia temática de su especialización, expresándolos de forma unívoca.

3. Se reordenaron alfabéticamente los términos, ya que su posición en la versión original no necesariamente era coincidente con la traducida.

4. Se revisaron los casos de sinonimia y homonimia, porque podía darse el caso de que diferentes términos en otro idioma se tradujeran por el mismo en español, o que

hubiera más sinónimos en español de los que expresaba el idioma extranjero, sin que existiera coincidencia con los posibles sinónimos en ese idioma.

En la actualidad se está utilizando en AENOR -el organismo que anteriormente se denominaba IRANOR- la versión del *TIT* de 1993 que, actualizada a 1996, recibieron en dicho organismo en julio de ese mismo año. El tesoro, recogido en formato de cinta magnética, se traduce simultáneamente con el uso de los descriptores necesarios para la indización.

El *Tesoro ISO* no es, en definitiva, un tesoro español, ni tan siquiera una edición en lengua española, ya que se emplean para el tratamiento de los documentos los mismos descriptores del tesoro original, el editado por AFNOR, aunque se esté procediendo a traducir los términos a medida que se van empleando en la descripción de documentos. El tiempo y dedicación que precisa la preparación de la traducción de un lenguaje documental son los motivos principales por los que, según los responsables del Servicio de documentación de AENOR, no se ha realizado hasta la fecha una versión española del citado tesoro.

#### **10. *Tesoro de Ciencias Farmacéuticas (en elaboración). (ICYT).***

Aunque se proyectó la elaboración de este tesoro en la temprana fecha de la que data el informe del ICYT, no se empezó a elaborar hasta 1995. En la actualidad no se está trabajando en el tesoro, de modo que no resulta posible calcular una posible fecha de finalización del mismo, según nos lo confirma Consuelo Ruíz, quien se responsabilizó del proyecto.

### 11. *Tesaurus de alimentación. Versión española del tesaurus de la CEE. (ICYT).*

El proyecto se redujo a la confección de un diccionario sobre la materia, de cuya preparación se hizo cargo el Grupo Termesp del ICYT, y que bajo la denominación *Alimentos: diccionario científico y tecnológico español-inglés-alemán-francés*, no se publicaría hasta 1994.

En resumen, de los once tesauros a los que hace referencia el informe del ICYT, sólo se puede confirmar la existencia de uno de creación española, el de Fundesco, y una versión española de un tesaurus internacional, el DIC.

Con objeto de recopilar otros datos que aportaran alguna luz sobre el tema acudimos a fuentes, más o menos autorizadas. Confrontados y analizados, dichos datos no siempre dieron resultados positivos, en muchos casos la esperanza de encontrar algún indicio de tesaurus nos llevó a encontrar léxicos, listas de encabezamientos de materias, etc., por ejemplo, investigamos la posible existencia de un proyecto de tesaurus de acuicultura que dió como resultado la existencia de un *Glosario de acuicultura*, que comenzó Manuela Vázquez en el seno del ICYT, en el que actualmente está trabajando Concha Ronda, y cuya edición verá la luz en fecha próxima.

De la información aportada en este capítulo se puede concluir que la creación de tesauros en España no ha sido prolífica históricamente. Concretamente, y en virtud de los datos contrastados, se puede afirmar la inexistencia de estos útiles hasta 1982, fecha en que empezaron a realizarse los primeros lenguajes documentales de términos

científicos y técnicos<sup>50</sup>. A partir de esa fecha el CSIC tomó la iniciativa de crear un programa de elaboración de tesauros especializados en distintas ramas del conocimiento. En 1986 el Instituto de Información y Documentación en Ciencias sociales y Humanidades (ISOC) programó la realización de tesauros sectoriales para indizar la documentación de las bases de datos creadas por dicha institución. También en el ICYT la Unidad de Investigación inició un proyecto de creación de tesauros especializados en los distintos campos científicos.

### **3.4. Tesauros elaborados en España**

En este apartado se ofrece información detallada de los tesauros construidos en nuestro país entre los años 1982 y 1997. Los datos relativos a cada uno de ellos se presentan organizados sistemáticamente por los siguientes campos: autor, edición, fecha, editor, lugar de edición, formato, número de ISBN, cobertura temática, formas de presentación y número de términos que los componen.

Como se podrá comprobar, en algunos casos se ha omitido la información de algunos campos porque no figuraba en los propios tesauros y, consultadas las fuentes relacionadas con los mismos, desconocían los datos. En este sentido, el campo más común sin cumplimentar es el de autor, debido a que se trata de trabajos realizados en colaboración.

---

<sup>50</sup>Las razones que argumentan Carmen Ortíz y Justo Fernández -especialistas en el área de traducción del ICYT- para explicar esta situación, es que es posible organizar la información con la Clasificación Taxonómica de Linneo en las áreas de botánica y zoología. Este sistema de clasificación, de posible uso internacional, que ha perdurado en el tiempo y que, al dividir y organizar la información en clase, subclase, orden, suborden, familia, subfamilia, género y especie, hace innecesaria la elaboración de tesauros en determinadas áreas de conocimiento.

## **1. Thesaurus: anàlisi metodològica i proposta de síntesi i per a l'Ajuntament de Barcelona.**

**Autor:** C. Udina

**Edición:** Primera

**Fecha:** 1987

**Editor:** Ajuntament de Barcelona

**Lugar de edición:** Barcelona

**Formato:** Impreso (copia de ordenador sin paginar)

**ISBN:** 84-505-6425-5

**Cobertura temática:** multidisciplinar

## **2. Tesauro Archivo**

**Autor:** Elisa Carolina de Santos Canalejo, con la colaboración del grupo de trabajo del Archivo General del Ministerio de Asuntos Exteriores.

**Edición:** Primera

**Fecha:** 1994

**Editor:** Ministerio de Asuntos Exteriores. Archivo general.

**Lugar de edición:** Madrid

**Formato:** Impreso (474 páginas) y en disquete

**ISBN:** 84-87661-43-2

**Cobertura temática:** Relaciones internacionales, política exterior, asuntos consulares, derechos humanos, economía, comercio, industria y energía, comunicaciones y obras públicas, etc.

**Formas de presentación:** sistemática, alfabética y permutada.

### **3. Thesaurus de Asuntos Sociales**

**Edición:** Segunda

**Fecha:** 1994

**Editor:** Ministerio de Asuntos Sociales, Centro de Publicaciones

**Lugar de edición:** Madrid

**Formato:** Impreso (297 páginas). Es posible su consulta en línea.

**ISBN:** 84-7850-050-2

**Cobertura temática:** Servicios sociales y protección social

**Formas de presentación:** Alfabética, sistemática y permutada

**Número de términos:** 1.710 descriptores y 429 no descriptores.

### **4. Tesauro de la base de datos de Fundesco**

**Autor:** Francisco Ortiz Chaparro

**Edición:** Primera

**Fecha:** 1987

**Editor:** Centro de Documentación de Fundesco

**Lugar de edición:** Madrid

**Formato:** Impreso (64 páginas)

**ISBN:** 84-84094-20-2

**Cobertura temática:** Tecnología de la información, electrónica, informática y telecomunicaciones.

**Formas de presentación:** Alfabética

Nota.- El tesauro no ha sido actualizado ni revisado desde su primera edición, lo cual lo hace inservible en la actualidad, algo que pasaría con cualquier otro lenguaje documental y que muy especialmente ocurre en una materia como las telecomunicaciones, que ha sufrido profundas modificaciones desde los años 80.

## **5. Thesaurus català d'educació**

**Autor:** Montserrat Gabarró i Parera [et al].

**Edició:** Primera

**Fecha:** 1992

**Editor:** Departament l'Ensenyament, Servei de Difusió i Edicions

**Lugar de edición:** [Barcelona]

**Formato:** Impreso (352 páginas), 2 disquetes y un cuadernillo.

**ISBN:** 84-393-2075-2.

**Cobertura temática:** Educación: recursos, equipamiento, aprendizaje, etc...

**Formas de presentación:** Índice alfabético, sin relaciones semánticas entre los términos, índice alfabético estructurado, permutado y parte sistemática representada gráficamente mediante terminogramas.

**Número de términos:** 3.462 descriptores y 1.000 no descriptores.

## **6. Thesaurus Ciències empresarials. Correspondencies en català, castellà i anglés.**

**Autor:** Francisca Buxó, Montserrat Catalan, Pepa Martínez, (coordinadoras)

**Edició:** Primera

**Fecha:** 1991

**Editor:** Fundació Empresarial Catalana y Escola Superior d' Administració y Direcció d'Empreses (ESADE).

**Lugar de edición:** Barcelona

**Formato:** Impreso (413 páginas)

**D.L.:** B-5143-1991

**Cobertura temática:** Temas relacionados con el campo empresarial.

**Formas de presentación:** jerárquica, permutada e índice alfabético, trilingüe

## **7. Tesauro de Ciencias de la información**

**Autor:** Antonio García Gutiérrez

**Edición:** Primera

**Fecha:** 1982

**Editor:** Centro de Documentación del Departamento de Documentación de la Facultad de Ciencias de la Información (IBERCOM)

**Lugar de edición:** Madrid

**Formato:** Impreso

**ISBN:**

**Cobertura temática:** Ciencias de la información

**Formas de presentación:** Alfabética, sistemática y permutada

**Número de términos:**

## **8. Tesauro de Defensa.**

**Autor:**

**Edición:** Segunda

**Fecha:** 1996

**Editor:** Ministerio de Defensa

**Lugar de edición:** Madrid

**Formato:** Impreso (283 páginas)

**ISBN:** 84-7823-453-5

**Cobertura temática:** Seguridad, Defensa y Fuerzas armadas

**Formas de presentación:** Alfabética, sistemática, índice auxiliar de nombres propios e índice auxiliar de topónimos.

**Número de términos:** unos 2000

## **9. Tesauro de Derecho**

**Autor:** Jorge Páez Maña y Pilar del Pozo Benéitez (coordinadores)

**Edición:** Primera

**Fecha:** 1997

**Editor:** CINDOC

**Lugar de edición:** Madrid

**Formato:** Impreso (2.241 páginas) [CD-Rom en preparación]

**ISBN:** 84-00-07682-6

**Cobertura temática:** Derecho, Documentación jurídica y terminología jurídica.

**Formas de presentación:** sistemática, alfabética y permutada

**Número de términos:** 19.494: 14.179 descriptores y 5.315 no descriptores.

## **10. Tesauro de drogodependencias**

**Autor:** Ricardo Mosquera Martínez, Francisco Rodríguez Martín, Inés Molowny Gómez.

**Edición:** Primera

**Fecha:** 1990

**Editor:** Ministerio de Sanidad y Consumo, Delegación del Gobierno para el Plan Nacional sobre Drogas.

**Lugar de edición:** Madrid

**Formato:** Impreso

**ISBN:** 84-7670-226-4

**Cobertura temática:** Drogodependencias

**Formas de presentación:** Alfabética y sistemática. Anexo con siglas, cuestionarios y tests utilizados en materia de drogodependencias.

## 11. Tesauro electrotécnico

**Autor:** Antonio García Gutiérrez (director), Elena Sigüenza Carbonell (coordinadora).

**Edición:** Primera

**Fecha:** 1993

**Editor:** Asociación de Investigación Industrial Eléctrica (ASINEL)

**Lugar de edición:** Móstoles (Madrid)

**Formato:** Impreso (4 volúmenes. V. I: 379 pp., V. II: 182 pp., V. III: 271 pp., V. IV: 151 pp.)

**ISBN:** 84-604-6878-X

**Cobertura temática:** El tesauro cubre el área electrotécnica con especial incidencia en los sectores que interesan a las empresas eléctricas españolas: fuentes de energía, equipos y materiales eléctricos, gestión de la explotación, etc...

**Formas de presentación:** Índice alfabético, índice jerárquico, índice permutado y anexo: diccionario inglés/español.

**Número de términos:** 7.764: 6.869 descriptores y 895 no descriptores.

## 12. Tesauro Geográfico administrativo EVETU

**Autor:**

**Edición:** Segunda

**Fecha:** 1993

**Editor:** Instituto Vasco de Administración Pública

**Lugar de edición:** [Oñate (Guipúzcoa)]

**Formato:** Impreso

**ISBN:** 84-7777-105-7

**Cobertura temática:** Denominaciones oficiales de las unidades administrativas de la

Comunidad Autónoma del País Vasco (territorios históricos, municipios, mancomunidades, parzonerías, comarcas y cuadrillas) y denominaciones oficiales de las Comunidades Autónomas del Estado y sus provincias así como de los Estados europeos.

**Formas de presentación:** Índice alfabético e índice jerárquico.

### **13. Tesoro de historia argentina**

**Autor:** Graciela G. Barcala de Moyano, Cristina Voena

**Edición:** Primera

**Fecha:** 1991

**Editor:** Anuario de Estudios Americanos

**Lugar de edición:** [Sevilla]

**Formato:** Impreso (40 páginas)

**ISBN:** Suplemento de Anuario de Estudios Americanos. Sección Historiografía y Bibliografía (1991) tomo XLVIII, nº 1, pp. 73-113.

**Cobertura temática:** Historia argentina

**Formas de presentación:** Alfabética, cronológica, sistemática y permutada.

**Número de términos:** 800

### **14. Tesaureus d'història contemporània de Catalunya**

**Autor:** Maribel Cuadrado Benítez [et al]

**Edición:** Primera

**Fecha:** 1996

**Editor:** Servei de Documentació d'Història Local de Catalunya. Departament d'Historia Moderna i Contemporània. Universitat Autònoma de Barcelona.

**Lugar de edición:** Barcelona

**Formato:** Impreso

**ISBN:** 84-7794-443-1

**Cobertura temática:** Historia contemporánea de Cataluña

**Formas de presentación:** alfabética y sistemática

**Número de términos:** 2.308 descriptores y 452 no descriptores

### **15. Thesaurus d'història social de la dona**

**Autor:** Montserrat Sebastià i Salat

**Edición:** Primera

**Fecha:** 1988

**Editor:** Comissió Interdepartamental de Promoció de la Dona, Generalitat de Catalunya.

**Lugar de edición:** Barcelona

**Formato:** Impreso ( 269 páginas)

**ISBN:** 84-393-0931-7

**Cobertura temática:** Ciencias sociales. Temática sobre mujer.

**Formas de presentación:** alfabética

### **16. Tesoro ICYT de biología animal**

**Autor:** Jesús Rey Rocha

**Edición:** Primera

**Fecha:** 1995

**Editor:** CINDOC

**Lugar de edición:** Madrid

**Formato:** Impreso (399 páginas), disquete y CD-Rom (es posible también su consulta en línea).

**ISBN:** 84-00-07510-2

**Cobertura temática:** Biología animal, Ciencias veterinarias, producción animal, etc...

**Formas de presentación:** alfabética, sistemática y permutada.

**Número de términos:** 3.781: 2.782 descriptores y 999 no descriptores.

### **17. Thesaurus de investigación educativa.**

**Autor:** Luis Lizasoain Hernández

**Edición:** Primera

**Fecha:** 1988

**Editor:** Editorial de la Universidad Complutense

**Lugar de edición:** Madrid

**Formato:** Impreso

**ISBN:** D.L. M 2161-1988

**Cobertura temática:** Materias relacionadas con los métodos y técnicas de la investigación educativa: evaluación, medición y técnicas de obtención de información o datos, etc...

**Formas de presentación:** Alfabética, sistemática y permutada. Incluye una lista inglés-castellano.

**Número de términos:** 1.423: 782 descriptores y 641 no descriptores.

### **18. Tesaurus ISOC de economía**

**Autor:**

**Edición:** Segunda/multilingüe

**Fecha:** 1995

**Editor:** Centro de Información y Documentación Científica

**Lugar de edición:** Madrid

**Formato:** Impreso (592 páginas)

**ISBN:** 84-00-07495-5

**Cobertura temática:** Economía

**Formas de presentación:** Alfabética, sistemática y permutada. Anexo: Identificadores y unidades monetarias.

**Número de términos:** unos 6.000

### **19. Tesauro ISOC de psicología**

**Autor:**

**Edición:** Segunda

**Fecha:** 1995

**Editor:** CINDOC

**Lugar de edición:** Madrid

**Formato:** Impreso (433 páginas). Es posible su consulta en línea.

**ISBN:** 84-00-07511-0

**Cobertura temática:** Psicología

**Formas de presentación:** Alfabética, sistemática y permutada más anexos con los nombres de tests, escalas y cuestionarios. El tesauro se completa con una lista onomástica de autores.

**Número de términos:** 2.788 descriptores, 1.612 no descriptores y 732 notas de alcance.

El anexo contiene 690 términos: 359 descriptores, 423 no descriptores y 331 notas de alcance.

## **20. Tesouro ISOC de topónimos**

**Autor:** José María Sánchez Nistal

**Edición:** Segunda

**Fecha:** 1996

**Editor:** CINDOC-CSIC

**Lugar de edición:** Madrid

**Formato:** Impreso (542 páginas) y versión informatizada

**ISBN:** 84-00-07576-5

**Cobertura temática:** Ciencias sociales. Topónimos político-administrativos.

**Formas de presentación:** Alfabética y sistemática

**Número de términos:** 26.000

## **21. Tesouro ISOC de urbanismo**

**Autor:**

**Edición:** Segunda

**Fecha:** 1992

**Editor:** Instituto Vasco de Administración Pública y CINDOC

**Lugar de edición:** Madrid

**Formato:** Impreso (470 páginas)

**ISBN:** 84-00-07260-X.

**Cobertura temática:** Ciencias sociales. Urbanismo y ordenación del territorio.

**Formas de presentación:** Alfabética, sistemática y permutada

**Número de términos:** 4.422: 3.652 descriptores y 770 no descriptores.

## **22. Tesoro de materias de la Base de datos INFOLUR**

**Autor:** Maite Tolosa (coordinadora)

**Edición:** Segunda

**Fecha:** 1994

**Editor:** Instituto Vasco de Administración Pública

**Lugar de edición:** Oñate (Guipúzcoa)

**Formato:** Impreso

**ISBN:** 84-7777-105-7

**Cobertura temática:** Urbanismo, ordenación territorial y medio ambiente.

**Formas de presentación:** alfabética y jerárquica.

**Número de términos:**

Nota.- El tesoro está editado conjuntamente con el Geográfico administrativo EVETU, y se incluyen en el Manual de la base de datos Infolur.

## **23. Tesoro de Medio Ambiente**

**Autor:**

**Edición:** Segunda

**Fecha:** 1990

**Editor:** Ministerio de Obras Públicas y Urbanismo

**Lugar de edición:** [Madrid]

**Formato:** Impreso (319 páginas) y soporte informático

**ISBN:** 84-7433-668-6

**Cobertura temática:** Medio ambiente, urbanismo, ordenación del territorio, temas ambientales.

**Formas de presentación:** Sistemática, alfabética y permutada

**Número de términos:** 2.320 descriptores y 608 no descriptores.

Nota.- Existen dos versiones posteriores editadas sólo en CD-Rom. La más reciente, correspondiente al año 1995, se compone de cerca de 3.000 descriptores y más de 700 no descriptores. En esta última versión existen numerosas modificaciones con respecto a la de 1990 (la última en papel) que afectan al contenido de los campos temáticos.

#### **24. Tesauro "Mujer"**

**Autor:**

**Edición:** Primera

**Fecha:** 1993

**Editor:** Instituto de la Mujer, Centro de Documentación

**Lugar de edición:** Madrid

**Formato:** Impreso (154 páginas)

**ISBN:** 84.7799-065-4.

**Cobertura temática:** Ciencias sociales. Temas relacionados con la mujer.

**Formas de presentación:** Alfabética, sistemática y permutada

**Número de términos:** 1.118: 714 descriptores y 404 no descriptores.

#### **25. Tesauro (Ministerio de relaciones con las Cortes y de la Secretaría del Gobierno).**

**Autor:**

**Edición:** [no está editado]

**Fecha:** 1989

**Editor:**

**Lugar de edición:**

**Formato:** Impreso

**ISBN:**

**Cobertura temática:** Ciencias sociales (temas jurídicos, constitucionales y políticos).

Recoge también nombres geográficos y de organismos.

**Formas de presentación:** alfabética y sistemática

## **26. Tesoro de patrimonio**

**Autor:** Antonio García Gutiérrez (coordinador)

**Edición:** Primera

**Fecha:** 1997

**Editor:** Consejería de Cultura del Instituto Andaluz del Patrimonio Histórico.

**Lugar de edición:** Sevilla

**Formato:** papel y CD-Rom.

**ISBN:** [En imprenta]

**Cobertura temática:** Abarca todos los temas relacionados con el Patrimonio histórico andaluz.

**Formas de presentación:** Alfabética, sistemática-alfabética y permutada.

**Número de términos:** 12.000 (aprox.)

Nota.- Se trata de un tesoro basado totalmente en la postcoordinación, para cuya elaboración se ha aplicado, como metodología básica, la gramática de casos.

## **27. Tesauro de puertos y costas**

**Autor:**

**Edición:** Primera (provisional)

**Fecha:** [1996]

**Editor:** Centro de Estudios y Experimentación de Obras Públicas, Centro de Información y Documentación Científica.

**Lugar de edición:** [Madrid]

**Formato:** Impreso (229 páginas)

**ISBN:**

**Cobertura temática:** Ingeniería de puertos, economía, comercio, transporte intermodal, etc.

**Formas de presentación:** alfabética, sistemática y permutada.

Nota.- Es un tesauro de uso interno para el Centro de Estudios y Experimentación de Obras Públicas (CEDEX), sólo puede utilizarse en otras instituciones previa autorización del Ente Público de Puertos del Estado.

## **28. Thesaurus de termes estadístics: correspondències en català, castellà i anglès.**

**Versión en español:** Thesaurus de términos estadísticos.

**Autor:**

**Edición:** Segunda

**Fecha:** 1993

**Editor:** Institut d'Estadística de Catalunya

**Lugar de edición:** Barcelona

**Formato:** Impreso (3 volúmenes)

**ISBN:** 84-393-2706-4

**Cobertura temática:** Estadística

**Formas de presentación:** alfabética, jerárquica y permutada.

**Número de términos:** 2.501 descriptores y 991 no descriptores.

**29. Thesaurus de términos relativos a relaciones raciales: with English equivalence: avec equivalence française: amb equivalència catalana.**

**Autor:** Ramón Soler y Fabregat

**Edición:** Primera

**Fecha:** 1996

**Editor:** CINDOC

**Lugar de edición:** Madrid

**Formato:** Impreso (X, 135 páginas)

**ISBN:** 84-00-07610-9

**Cobertura temática:** Ciencias sociales. Relaciones raciales.

**Formas de presentación:** Índice sistemático, índice alfabético estructurado e índice alfabético permutado.

**Número de términos:** 1.434: 1.023 descriptores y 402 no descriptores.

**30. Thesaurus de la Universitat de Barcelona: BUB, 1992.**

**Autor:** Carme Cambrodí amb la col. de M<sup>a</sup> Teresa Tarrida

**Edición:** Primera

**Fecha:** 1992

**Editor:** Area de Procés Tècnic de la Biblioteca de la Universitat de Barcelona

**Lugar de edición:** Barcelona

**Formato:** Impreso (459 páginas) y soporte informático (Consultable en línea)

**ISBN:** 84-7875-864-X

**Cobertura temática:** Multidisciplinar

**Formas de presentación:** Alfabética y permutada

**Términos:** 8.500

En el catálogo de la Biblioteca de la Universitat de Barcelona figuraban las siguientes referencias: **Thesaurus: Anàlisi metodològica i proposta de síntesi per a l'Ajuntament de Barcelona** y **Thesaurus semàntic interdimensional: selecció dels classificadors més habituals en la dinàmica municipal**, éste último editado por el Ajuntament de Barcelona. No nos fue posible analizar ninguno de ellos debido a que el único centro en el que parecen existir ejemplares es el Instituto Municipal de Animación y Esplai (IMAE), que ha pasado a denominarse Agencia Municipal de Servicios para las Asociaciones, y que se encuentra actualmente en proceso de traslado, no pudiéndose consultar sus fondos.

#### **3.4.1. Tesoros en fase de elaboración**

A continuación se señalan los tesauros que se encuentran en fase de elaboración, de los que hemos tenido conocimiento. Estos lenguajes abarcan áreas de conocimiento diversas, tales como artes escénicas, derecho, obras públicas, ciencia de los materiales y sociología.

- **Tesauro sobre artes escénicas.** Constituye un lenguaje ambivalente que cubre la terminología de las artes escénicas en general, cuya publicación, a cargo del Instituto del teatro de Barcelona, está prevista para finales del presente año. Trata de cubrir la documentación existente en la biblioteca de dicho teatro y de su archivo, que asciende a un total aproximado de 120.000 volúmenes.

- **Tesauro jurídico** (Facultad de Derecho de Sevilla). Se trata del expurgo de la cuarta edición (1995) del denominado *Thesaurus de la Biblioteca de la Universidad de Derecho con notaciones de la CDU*, que consiste en una lista de encabezamientos de materias compilada por Arcadio Castillejo Benavente y Carmen Jiménez-Castellanos Ballesteros, que pasará a estructurarse como tesauro.

- **Tesauro de Obras hidráulicas.** Se está elaborando en el Centro de Estudios y Experimentación de Obras Públicas (CEDEX). La fecha aproximada de publicación es septiembre del presente año.

- **Tesauro de Ciencia de los materiales.** Constituirá la versión española del tesauro de términos metalúrgicos *Thesaurus of the American Society of materials*. Se encuentra en fase de revisión previa a la publicación, y lo han llevado a cabo Amelia de Irazazábal y Sofía Álvarez.

- **Tesauro de Sociología.** Realizado por Angel Villagrà Rubio, se encuentra en la fase final de su construcción, a falta de la inclusión de algunas relaciones asociativas. La fecha estimada para su finalización es el primer trimestre de 1998. El ISOC se hará cargo de su publicación. El tesauro, que se realiza simultáneamente en las versiones española, inglesa y francesa, constará de un total aproximado de 4.000 términos, organi-

zados en tres formas de presentación: alfabética, sistemática y permutada, y estará disponible en papel y en disquete.

- **Tesoro del CIS (Centro de Investigaciones Sociológicas)**. Basado en material recogido en las encuestas sobre opinión pública, realizadas durante los últimos veinticinco años. Constará de 3.000 términos aproximadamente y se organizará en las tres formas de presentación tradicionales: alfabética, sistemática y permutada.

#### **3.4.2. Tesoros internacionales con versión o traducción en español, catalán o valenciano:**

Se trata de tesauros adaptados o traducidos al español, catalán o valenciano, de uso regular para el almacenamiento y recuperación de información en bases de datos de diferentes centros de documentación o bibliotecas especializadas.

Al igual que en el epígrafe 3.4., la información de este apartado se organiza en base a los datos de autor, edición, fecha, etc...

#### **Tesoro AGROVOC: Tesoro agrícola multilingüe**

**Autor:**

**Edición:** Segunda, multilingüe

**Fecha:** 1992

**Editor:** Apimondia, por acuerdo con la Organización de las Naciones Unidas para la Agricultura y la Alimentación.

**Lugar de edición:** Roma

**Formato:** Impreso (2 tomos: 1º t.: 855 págs., 2º t.: 305 págs.)

**ISBN:** 88-7643-009-1

**Cobertura temática:** Agricultura en general, geografía e historia, economía, desarrollo agrícola y sociología rural, recursos naturales y medio ambiente, etc.

**Formas de presentación:** Índice alfabético que comprende tres categorías de entradas: los descriptores, los no descriptores y las formas permutadas correspondientes a descriptores y no descriptores representados por expresiones compuestas. Se completa con un índice alfabético de cada una de las términos del tesoro en español, inglés y francés.

**Número de términos:** 14.714 descriptores, 8.495, 11.048 y 7.602 no descriptores en las versiones inglesa, española y francesa, respectivamente.

### **Tesoro de la Biblioteca de la CCE (basado en el microtesoro de la OCDE)**

**Autor:**

**Edición:**

**Fecha:** 1988

**Editor:** ICYT

**Lugar de edición:** Madrid

**Formato:** Impreso (260 páginas)

**ISBN:**

**Cobertura temática:** Disciplinas científicas múltiples

**Formas de presentación:**

**Número de términos:**

Nota.- En 1996 los alumnos de la Universidad Pontificia de Comillas, de la asignatura Terminología, de la Facultad de Traducción e Interpretación, hicieron la versión española actualizada de este tesoro, bajo la dirección de la profesora Amelia de Irazazábal, estando en fase de revisión previa a su publicación.

### **Tesauro de carreteras**

**Autor:** Concepción Lallana del Valle (comp.)

**Edición:** Primera

**Fecha:** 1993

**Editor:** Centro de Estudios y Experimentación de Obras Públicas, Gabinete de Formación y Documentación.

**Lugar de edición:** Madrid

**Formato:** Impreso (454 páginas) y disquete

**ISBN:** 84-7790-165-1

**Cobertura temática:** Carretera y transporte por carretera

**Formas de presentación:** Índice numérico (en este índice se encuentran todos los términos incluidos en el tesauro -descriptores y no descriptores-, agrupados en las 50 áreas temáticas de que consta), índice alfabético, índice de sinónimos e índice permutado.

**Número de términos:** 4.453

**Nota.-** Se trata de la versión española del Tesauro DIC, pero, a diferencia de éste, la parte sistemática no se representa mediante diagramas de flechas.

### **Tesauro EUDISED: Thesaurus multilingüe para el tratamiento de la información relativa a la educación.**

**Autor:** preparado por Jean Viet

**Edición:** Primera española

**Fecha:** 1975

**Editor:** Mouton & Consejo de Europa

**Lugar de edición:** La Haya

**Formato:** Impreso (un volumen por versión lingüística)

**ISBN:**

**Cobertura temática:** Ciencias de la educación. Incluye nombres de países, regiones y de organizaciones internacionales.

**Formas de presentación:** sistemática y alfabética

### **Tesaurus EUROVOC**

Versión valenciana del tesaurus de las Comunidades europeas, editado en Bruselas y Luxemburgo.

**Edición:** Segunda

**Fecha:** 1993

**Editor:** Conselleria d'Administració Pública

**Lugar de edición:** Valencia

**Formato:** Impreso (3 tomos. I: 268 páginas; II: 146 páginas; III: 153 páginas).

**ISBN:** 84-482-0344-5

**Cobertura temática:** Vida política, relaciones internacionales, asuntos sociales, derecho, educación, comunicación, industria y energía, sector agroalimentario, etc.

**Formas de presentación:** alfabética, sistemática y permutada.

**Número de términos:** descriptores: 5.359, no descriptores: 5.170.

### **Thesaurus de la Formación Profesional. CEDEFOP**

**Edición:** Primera española

**Fecha:** 1990

**Editor:** Ministerio de Trabajo y Seguridad Social

**Lugar de edición:** Madrid

**Formato:** Impreso (222 páginas)

**NIPO:** 210-90-011-X

**Cobertura temática:** Formación profesional

**Formas de presentación:** Alfabética, permutada y lista de "top terms".

**Thesaurus Iconogràfic: Sistema descriptiu de les representacions.**

Nota.- Se trata de la adaptación al catalán del Thesaurus iconographique realizado por Garnier<sup>51</sup>, cuyos datos no podemos incluir en este anexo por no habernos sido posible conseguir un ejemplar del mismo.

**Tesouro OIT: Terminología de trabajo, empleo y formación**

**Autor:**

**Edición:** Cuarta, multilingüe

**Fecha:** 1992

**Editor:** Oficina Internacional del Trabajo

**Lugar de edición:** Ginebra

**Formato:** Impreso (557 páginas)

**ISBN:** 92-2-006429-4

**Cobertura temática:** Desarrollo económico y social, derecho, ciencias sociales, educación, actividades económicas, trabajo y empleo, población, relaciones raciales y migración, etc.

**Formas de presentación:** Índices alfabéticos en inglés, francés y español, que remiten

---

<sup>51</sup>GARNIER, F. *Thesaurus iconographique: système descriptif des représentations*. Paris: Le Léopard d'or, 1981.

al usuario a la presentación sistemática, en la que se presentan los descriptores provistos de sus relaciones semánticas.

**Número de términos:** Descriptores: 4.039, no descriptores: 1.019 en inglés, 696 en francés y 603 en español.

**Tesaurus de Popin: Tesaurus multilingüe sobre población**

**Autor:** Preparado por Jean Viet

**Edición:** Segunda

**Fecha:** [1984]

**Editor:** Fondo de las Naciones Unidas para las actividades en materia de población (FNUAP)

**Lugar de edición:** Nueva York

**Formato:** Impreso

**ISBN:**

**Cobertura temática:** materias relacionadas con la población (demografía, migración, mortalidad, sanidad, etc...)

**Formas de presentación:** alfabética, sistemática y permutada

**Número de términos:**

**Tesaurus SPINES**

**Autor:**

**Edición:** Primera española

**Fecha:** 1988

**Editor:** UNESCO/ICYT

**Lugar de edición:** Madrid

**Formato:** Impreso (2 v., 1º v.: V + 358 páginas; 2º v.: LI + 716 páginas).

**ISBN:** 84-00-06802-5

**Cobertura temática:** Disciplinas científicas múltiples

**Formas de presentación:** Lista alfabética estructurada, índice temático de descriptores, índice de palabras significativas contenidas en los descriptores, vocabulario inglés-español, vocabulario francés-español, índice de organizaciones y programas internacionales.

**Número de términos:** 10.832: 7.993 descriptores y 2.839 no descriptores.

**Tesoro de términos para ciencias acuáticas y pesqueras (traducción del Thesaurus of terms for aquatic sciences and fisheries).**

**Autor:** (Traducido por Alicia Teresa Mileo)

**Edición:** Primera

**Fecha:** 1985

**Editor:** Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) y la Pontificia Universidad Católica de Chile.

**Lugar de edición:** Roma

**Formato:** Impreso

**ISBN:** 92-5-302233-7

**Cobertura temática:** Ciencias acuáticas y pesqueras

**Formas de presentación:** Alfabética y lista de términos categorizados por facetas.

**Número de términos:** 1900 descriptores y 950 no descriptores

### **Tesaurus de términos de comercio internacional**

**Autor:**

**Edición:** Segunda

**Fecha:** 1985

**Editor:** Centro de Comercio Internacional UNCTAD/GATT

**Lugar de edición:** Ginebra

**Formato:** Impreso (146 páginas)

**ISBN:** 53-109568-0

**Cobertura temática:** Comercio internacional y promoción comercial, tomando en consideración particularmente a los países en desarrollo.

**Formas de presentación:** Lista alfabética, lista numérica.

**Número de términos:**

### **Thesaurus Textil Internacional**

**Autor:**

**Edición:** Primera

**Fecha:** 1971

**Editor:** Departamento de Documentación de la Escuela Técnica Superior de Ingenieros Industriales.

**Lugar de edición:** Tarrasa

**Formato:** Impreso

**Cobertura temática:** Industria textil

**Formas de presentación:** alfabética

Nota.- Constituye la edición española de Thesaurus del Instituto de Tecnología de Massachusetts.

**Macrothesaurus para el tratamiento de la información relativa al desarrollo económico y social.**

**Autor:** Edición preparada por Jean Viet

**Edición:** Cuarta

**Fecha:** 1991

**Editor:** Organización de Cooperación y Desarrollo Económico (OCDE)

**Lugar de edición:** Paris

**Formato:** Impreso (428 páginas)

**ISBN:** 92-64-31882-8

**Cobertura temática:** Economía, sociología

**Formas de presentación:** Alfabética, jerárquica y permutada

**Tesaurus de la UNESCO**

**Autor:** Jean Aitchison

**Edición:** Primera

**Fecha:** 1984

**Editor:** UNESCO

**Lugar de edición:** Paris

**Formato:** Impreso (2 v., 1º v: 419 páginas; 2º v.: 615 páginas).

**ISBN:** 92-3-301469-6

**Cobertura temática:** Educación y ciencia, ciencias sociales, cultura, comunicación y documentación.

**Formas de presentación:** Sistemática, alfabética y permutada. Tiene además una lista de identificadores, aunque no se publica.

**Número de términos:** Unos 8.500.

Nota.- No se trata de una versión española, sino de su simple traducción al castellano, realizada por Jean Viet.

## Capítulo 4

### La recuperación de información

## 4.1. Introducción

Como ya hemos dicho anteriormente, para buscar información en un fondo documental es posible utilizar el propio lenguaje de los documentos almacenados o emplear un lenguaje controlado que evite la ambigüedad característica del natural. En ambos casos se puede recurrir a técnicas que agilizan el proceso y que exponemos en el presente capítulo.

La recuperación de información es la última fase del tratamiento documental, y aquella en la que se da respuesta a las solicitudes de información de los usuarios.

La recuperación puede ser manual o automática. Van Slype denomina a esta última *recuperación asistida por el ordenador*<sup>1</sup>, que a nosotros nos parece más adecuada, porque entendemos que, ya sea llevada a cabo la operación por un documentalista o por el propio usuario (como viene siendo habitual con el desarrollo de sistemas informáticos de fácil acceso), la recuperación no es completamente automática.

El desarrollo que han experimentado los sistemas de recuperación automatizados en los últimos treinta años, ha acelerado el reemplazo de métodos manuales de recuperación por distintos medios de búsqueda automatizados.

---

<sup>1</sup>SLYPE, G.v. **Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales**. Madrid: Fundación Germán Sánchez Ruipérez, 1991, p. 174.

La recuperación de información manual o asistida por ordenador es tarea que *concierna a la representación, almacenamiento, organización y acceso a los documentos*<sup>2</sup>, y por tanto, la calidad de la misma depende de la calidad de la indización y de la fase de recuperación. A este respecto escribe García Marco que *la búsqueda depende de las habilidades y estrategias del usuario, el entorno físico y tecnológico, la estructura de los datos, el sistema de representación de los conceptos (lenguajes de indización y clasificación) y el lenguaje de recuperación*<sup>3</sup>.

Tanto la evaluación de la recuperación como la del lenguaje de recuperación implican la valoración de la calidad del propio Sistema de Información, cuyas características describimos en el siguiente epígrafe.

#### **4.2. Sistemas de Información**

Un Sistema de Información es, según Miguel Castaño y Piattini Velthuis *un conjunto de elementos ordenadamente relacionados entre sí de acuerdo con unas ciertas reglas que aporte el sistema objeto (es decir, la organización a la cual sirve y que le marca las directrices de funcionamiento) la información necesaria para el cumplimiento de sus fines, para lo cual tendrá que recoger, procesar y almacenar datos, procedentes tanto de la misma organización como de fuentes externas, facilitando la recuperación, elaboración y presentación de los mismos.*<sup>4</sup>

---

<sup>2</sup>SALTON, G. & MCGILL, M.J. **Introduction to modern information retrieval**. New York, etc.: McGraw Hill Publishing Company, 1983, p. 1.

<sup>3</sup>GARCÍA MARCO, F.J. De la consulta de catálogos a la gestión de información: Tensiones hacia el cambio en el diseño de OPACs. **Boletín de la ANABAD**, v. 41, nº 3-4 (1991) pp. 325-334.

<sup>4</sup>MIGUEL CASTAÑO, A. de & PIATTINI VELTHUIS, M.G. **Concepción y diseño de bases de datos: Del modelo E/R al modelo relacional**. Madrid: RA-MA, 1993, p. 17.

Los Sistemas de Información (SI) pueden ser manuales o pueden estar mecanizados, aunque en la actualidad, prácticamente todo SI se apoya en técnicas informáticas, realizándose el tratamiento y recuperación de la información, por regla general, mediante sistemas de gestión de bases de datos.

Un sistema de información se compone de: los datos; el equipo informático, tanto el equipamiento físico (los ficheros donde se almacenan los datos) como los programas (que gestionan toda la aplicación); un gestor de los datos y los usuarios. No obstante, contando con los mismos componentes, cada sistema ofrece características propias.

Miguel Castaño y Piattini Velthuis consideran que *los datos pueden ser formateados o no formateados, los textuales, propios de los sistemas documentales, son no formateados. Atendiendo a estos aspectos, se suele distinguir entre dos tipos de sistemas de tratamiento:*

- *Sistemas de Recuperación de Información (SRI): se dedican al tratamiento de datos no formateados (documentos). Estos sistemas proporcionan facilidades de thesaurus, búsqueda en texto libre, etc.*

- *Sistemas de Gestión de Bases de Datos (SGBD): se dedican al tratamiento de datos formateados.*<sup>5</sup>

Las bibliotecas constituyen los primeros SRI que, con sus catálogos, sumarios, boletines de resúmenes, etc., realizaban y aún siguen realizando, las tareas de recuperación de información mediante un sistema de búsqueda lineal y acumulativa. En

---

<sup>5</sup> MIGUEL CASTAÑO, A. de & PIATTINI VELTHUIS, M.G. op. cit., p. 19.

los modernos SRI es posible establecer relaciones semánticas previamente al proceso de búsqueda, combinando el número de conceptos necesarios para seleccionar la información relevante.

Un SRI se compone de los elementos propios de todo SI, mencionados más arriba, y de elementos externos. Los elementos externos son los lenguajes documentales, que facilitan la normalización del proceso de recuperación; los recursos sintácticos, que se pueden emplear para formular las estrategias de búsqueda; el interfaz, que permite la comunicación hombre-máquina; y los registros resultantes del proceso de búsqueda.

De los lenguajes documentales hemos tratado en el capítulo segundo de esta tesis doctoral, y más concretamente del tesoro, de uso generalizado en SRI por sus prestaciones en áreas temáticas especializadas. Los registros son unidades que recogen la descripción de cada documento. Están compuestos por subunidades denominadas campos, que contienen un único dato descriptivo del documento. De los campos que componen cada registro hablaremos detalladamente en el apartado dedicado a las bases de datos.

En cuanto a los recursos sintácticos que ofrecen los SRI, nos limitamos en este epígrafe a citar algunos de los más empleados, ya que, necesariamente hablaremos de sus características al tratar de los procedimientos de búsqueda documental, tema que se desarrolla en el epígrafe 4.4.4. de este capítulo. Los recursos sintácticos más comúnmente utilizados en la recuperación de información son: el truncamiento, que permite aumentar la exhaustividad de la búsqueda; los operadores booleanos, los enlaces de proximidad y la búsqueda mediante ponderación de los descriptores.

El cuarto elemento externo que compone un SRI es el interfaz, que es la parte del sistema de recuperación que posibilita la interacción con el usuario. García Marco escribe que *en una biblioteca el interfaz de usuario lo compone la parte pública del OPAC, pero también el tiempo y cualidades dedicadas en los usuarios por parte del personal bibliotecario*<sup>6</sup>.

Para hacer una búsqueda hay que utilizar comandos y operadores con los términos adecuados para construir una estrategia que produzca resultados acordes con la pregunta formulada. Los comandos que empleaban las primeras bases de datos exigían, por su complejidad, su aprendizaje previo para poder manejar la base. Si se tiene en cuenta que cada servicio utiliza comandos o lenguajes de recuperación diferentes, se puede comprender el esfuerzo del usuario a la hora de consultar los SRI. Existe abundante literatura sobre las ventajas que aportaría el empleo de un lenguaje de interrogación común, pero, hasta la fecha, no se ha logrado este objetivo.

Una alternativa a los comandos es el uso de menús, de fácil manejo e innecesario aprendizaje. El usuario tiene únicamente que ir seleccionando los pasos, las opciones que le sugiere la pantalla en forma de menú. El programa se encarga de convertir las cuestiones formuladas por el usuario a los comandos y conecta los términos de la búsqueda con los operadores adecuados.

García Marco<sup>7</sup>, en el trabajo que citamos anteriormente, realiza una introducción histórica del interfaz hombre-máquina, en la que hace referencia a la evolución desde la

---

<sup>6</sup>GARCÍA MARCO, F.J. *Interfaces amigables para la recuperación de la información bibliográfica*. Scire, v. 1, nº 1 (1995) p. 129.

<sup>7</sup>GARCÍA MARCO, F.J. *Op. cit.* pp. 132-135.

comunicación mediante las referidas complejas instrucciones hasta el desarrollo de menús, fácilmente manejables para navegar por la base de datos.

A este respecto, Foskett hace referencia al significativo desarrollo en "microcomputer" que constituyó la introducción de Apple Macintosh en 1984. *Con anterioridad - escribe- era necesario teclear y conocer los comandos de memoria o consultando un manual. Apple Macintosh fue el primer microordenador en usar un interfaz de usuario gráfico (GUI) en el que los comandos estaban representados en la pantalla por iconos, que se seleccionaban por medio de un "pointing" recurso, como un ratón. El ratón se utiliza para mover el cursor hacia la posición deseada en la pantalla. Alternativamente, menús -listas de comandos- se podían seleccionar de la misma forma. La necesidad de teclear los comandos se redujo al mínimo, simplificando las tareas. Un GUI de equivalente calidad para IBM, compatible con PCs tuvo que esperar hasta Windows 3.1 en 1991<sup>8</sup>.*

Pero la alternativa de los menús no está exenta de desventajas. El empleo de estos recursos resulta más lento que el de comandos directos, especialmente en el supuesto de que el usuario necesite consultar pantallas de ayuda, que suelen complicar la tarea, en particular cuando el consultante no está acostumbrado a trabajar solo con el sistema. Por otra parte, la independencia que se otorga al usuario para manejar la base conlleva la desventaja de su aislamiento y la ausencia de ayuda, por parte del personal especializado, para seleccionar los términos adecuados de la búsqueda, así como la preparación o modificación de las estrategias.

---

<sup>8</sup>FOSKETT, A.C. **The subject approach to information**. 5ª ed. London: Clive Bingley, 1996, p. 45.

Tanto si se utilizan comandos o menús es determinante para el éxito de la comunicación hombre-máquina la implantación, en los distintos SRI, de interfaces que, con determinadas características faciliten el manejo del sistema. Un interfaz que reúna las cualidades de amigabilidad, economía, ergonomía e interactividad, ampliará la capacidad de comunicación y permitirá realizar las tareas de búsqueda documental en las mejores condiciones de seguridad, comodidad y economía para el usuario. El factor interactividad posibilita el diálogo hombre-máquina a lo largo de todo el proceso de búsqueda y da la oportunidad al usuario de ir modificándola a medida que progresa, de tal manera que cada documento o cada dato encontrado influya en la búsqueda siguiente. *Tales búsquedas, -dice Foskett- en las que el curso de los acontecimientos se modifica continuamente a la luz de la información que se va obteniendo, pueden ser descritas como heurísticas*<sup>9</sup>.

La posibilidad que ofrecen las búsquedas heurísticas de modificar continuamente las estrategias, no existen cuando al frente del proceso está un bibliotecario/documentalista; sí que es posible, sin embargo, que dicho especialista muestre al usuario los resultados de una primera búsqueda, pudiendo éste modificar la estrategia teniendo en cuenta los documentos relevantes obtenidos. Puede realizar entonces un segundo intento y repetir el proceso hasta que el usuario quede satisfecho con los resultados. A este tipo de búsqueda que no se modifica de forma continua sino a intervalos Foskett la denomina *búsqueda interactiva*<sup>10</sup>.

En cualquier caso, para que el usuario pueda elegir entre cualquiera de las opciones de búsqueda, la pantalla debe ofrecer en todo momento alternativas para realizar

---

<sup>9</sup>FOSKETT, A.C. Op. cit. p. 26.

<sup>10</sup>FOSKETT, A.C. Op. cit., p. 26.

la consulta sin necesidad de ayuda experta. Se trata de dar autonomía a usuarios más o menos experimentados en el manejo de sistemas de información. El interfaz debe ir indicando todos los pasos a seguir para evitar que el consultante tenga que memorizar los símbolos o elementos que lo componen (como ocurre, en algunos casos, con los comandos). La agrupación de dichos elementos en recuadros facilita el acceso a la base y le da al usuario la seguridad de que sabe controlar el sistema sin estar pendiente de la ayuda externa.

Antes de terminar este apartado queremos señalar que, además de las facilidades suministradas por los lenguajes y los recursos señalados, los distintos SRI proporcionan otros medios suplementarios para ayudar a los usuarios a obtener el máximo provecho de los datos contenidos en la misma. Se trata de un conjunto de programas o procedimientos para la carga de ficheros, generadores de listados o tablas, etc.

### **4.3. Bases de datos**

Las bases de datos son los instrumentos que emplean los Servicios de Recuperación de Información para cumplir las funciones de preparación y recuperación de los datos.

*Una base de datos es una colección o depósito de datos integrados, con redundancia controlada y con una estructura que refleje las interrelaciones y restricciones existentes en el mundo real; los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de éstas, y su definición y descripción, únicas para cada tipo de datos, han de estar almacenadas junto con los mismos. Los procedimientos de actualización y recuperación, comunes y bien determinados, habrán*

*de ser capaces de conservar la integridad, seguridad y confidencialidad del conjunto de los datos.*<sup>11</sup>

En función del tipo de información que sirve, una base de datos puede ser factual, si recoge datos numéricos; bibliográfica, si fundamenta su actividad en la oferta de citas tanto de libros como de publicaciones periódicas; o puede tratarse de bases que limitan su campo a informar al usuario acerca de patentes, direcciones, etc. Además de este tipo de información, las bases de datos pueden tratar temas multidisciplinarios o especializarse en un área de conocimiento en particular.

Vamos a referirnos aquí a las bases de datos bibliográficas -el tipo de base que consultamos para realizar la parte nuclear de esta investigación-, que recogen la producción editorial sobre cualquier rama del saber o sobre un área temática en particular.

Bares denomina a éstas, *bases documentales*<sup>12</sup>, las cuales define como *conjunto de ficheros que se refieren al mismo campo de interés, y que forman un todo organizado con respecto al logical documental que les soporta.*

En este tipo de bases es necesario recurrir a la técnica del fichero inverso, cuya organización particular hace posible la selección de un fondo documental, a partir de descriptores o de palabras clave. En estos ficheros, los descriptores van seguidos de las referencias en la que aparecen. En el momento de la búsqueda se introducen dichos descriptores o palabras clave en la memoria; las listas correspondientes a cada uno de

---

<sup>11</sup>MIGUEL CASTAÑO, A. & PIATTINI VELTHUIS, M.G. Op. cit., p. 46.

<sup>12</sup>BARES, M. *La búsqueda documental en el contexto telemático: Modalidades de automatización y utilización de bases de datos.* Madrid: Díaz de Santos, 1988, p. 104.

ellos se localizan en el fichero inverso. El ordenador puede realizar rápidamente la intersección lógica de diferentes listas para encontrar las referencias comunes. *En comparación con una búsqueda secuencial (se barre todo) que se refiere al conjunto, esta técnica restringe considerablemente el campo a explorar y trae consigo una ganancia de tiempo considerable*<sup>13</sup>.

Una base de datos bibliográfica puede ser referencial o textual. Las referenciales son aquellas que ofrecen la información necesaria para acceder a determinado documento, pero no el documento en sí. Algunas bases de este tipo contienen un adelanto del contenido de los documentos porque incluyen en sus registros un resumen de aquellos, pero, en cualquier caso, el lector no tiene acceso al documento fuente. Las bases de datos textuales proporcionan el texto completo, no sólo la referencia del original, de tal manera que el usuario no precisa localizar otras fuentes para satisfacer sus necesidades de información.

#### **4.3.1. Registros de una base de datos**

Las bases de datos bibliográficas se componen de registros que describen las características físicas y de contenido del documento. Cada uno de estos registros está constituido por una serie de campos, cuyo número y tipología puede variar de una base a otra pero que, en general, ofrecen la información básica o complementaria que permite la recuperación de determinada obra mediante la consulta de todos o solamente de alguna de las unidades componentes de dicho registro. Los campos que comprende un registro son los siguientes:

---

<sup>13</sup>BARES, M. Op. cit. p. 115.

**a) Código de acceso a la base de datos**

Se trata de un código numérico o alfanumérico que se asigna al registro en el momento de su incorporación a la base.

**b) Autor**

Campo que refleja la mención de responsabilidad de la obra, ya se trate de autor personal, corporativo, etc.

**c) Título**

Transcripción literal del título del documento. Algunas bases incorporan un campo para título original, teniendo en cuenta que existe la tendencia a traducir al inglés el título de obras escritas en cualquier idioma, como medio de normalización.

**d) Fecha de publicación**

Campo que delimita el año de publicación de los documentos.

**e) Forma de publicación**

Se refiere al tipo de documento de que se trate: folletos, libros, revistas, actas de congresos, etc.

**f) ISBN o ISSN**

Siglas correspondientes al Número Internacional Normalizado de registro de Libros o de Publicaciones seriadas, respectivamente.

**g) Fuente**

Este campo facilita la información relativa al origen editorial del documento. En función del tipo de obra indica el título de la revista, título de la obra completa, en el caso de tratarse del capítulo de una obra referenciado independientemente, nombre genérico del congreso, en el caso de ponencias, o nombre de la editorial, cuando la obra es una monografía.

Cualquiera de estas referencias permite localizar el documento fuente, en el supuesto de que fuera necesaria su consulta.

**h) Lengua original del documento**

Expresa el idioma en el que están escritas y publicadas las obras.

**i) Clasificación**

Incluye la notación numérica o alfanumérica correspondiente al documento clasificado. Esta codificación puede o no ser coincidente con la localización de la obra en el fondo documental.

El sistema de clasificación más generalizado en nuestro país, y el que adoptan la mayoría de las bases de datos es la *Clasificación Decimal Universal (CDU)*, que permite acoplar la organización del fondo y el catálogo de la biblioteca.

El campo de clasificación facilita la búsqueda en bases de datos con amplio volumen de documentos, así como en aquellas de carácter interdisciplinar, porque permite restringir la búsqueda a un área temática en particular.

#### **j) Descriptores**

Se recogen en este campo los puntos de acceso temáticos. Puede tratarse de descriptores extraídos de un tesoro, o de palabras clave tomadas del propio texto. En ambos casos la selección la realiza el analista en el momento de la indización, con el propósito de representar el contenido conceptual del documento.

El número de términos que incluyen las diferentes bases en este campo fluctúa considerablemente, según la política de indización seguida en cada centro.

#### **k) Resumen**

Incluye resumen indicativo o informativo del documento. Constituye un campo de gran ayuda para el consultante ya que en él se ofrece información sobre objetivos, metodología, resultados, conclusiones y recomendaciones del documento de referencia. Estos datos permiten mejorar la búsqueda cuando se trabaja en modo interactivo, y son insustituibles cuando la ecuación de búsqueda se prepara en lenguaje natural.

La consulta de este campo suele resultar suficiente para conocer las características esenciales de una obra, ahorrando tiempo y dinero al usuario, que no necesita acudir al documento original para satisfacer sus necesidades de información, especialmente si el tipo de resumen que incluye la base es informativo.

La importancia de la incorporación del resumen para poder valorar la relevancia de los documentos recuperados ha sido muy grande en este trabajo de investigación, porque tuvimos que analizar un total de 25.399 referencias, para conocer sus índices de relevancia. La labor, que si fue muy laboriosa, se multiplicó en aquellas bases en las que no existía resumen de los documentos, dado que, en muchos casos, fue necesaria la consulta de los obras originales para poder realizar una valoración objetiva de sus contenidos.

*Como dice Moya Anegón, en la mayoría de los sistemas de recuperación de información existen cierto tipo de campos que aunque no se pueden utilizar como términos de búsqueda en las operaciones de recuperación, se pueden utilizar como limitadores de conjuntos de documentos seleccionados previamente. Estos campos se llaman de esta forma porque su función es la de reducir el tamaño de los conjuntos de documentos obtenidos mediante la utilización de los términos indexados y los operadores que los pueden relacionar. Este tipo de operación de búsqueda en los sistemas de gestión de catálogos son especialmente útiles para los usuarios porque les permite acceder a información bibliográfica por códigos y números no indexados, lo que reduce el número de índices y amplía las posibilidades de recuperación..... Las informaciones que de forma más usual han sido objeto de este tipo de tratamiento son: el código de tipo de material, los códigos de país y lengua de la publicación, y las fechas, estas últimas sujetas a un tratamiento especial, pues la limitación aquí consiste normalmente en uno de entre cuatro*

*tipos de operaciones aritméticas, igual que (=), mayor que (>), menor que (<) y entre (> <)*<sup>14</sup>.

### **Ejemplos de registros**

Todas las bases de datos consultadas para nuestro trabajo experimental son bibliográficas, pero la estructura de sus registros no es idéntica. A continuación mostramos un ejemplo de registro de cada una ellas.

#### **Base de datos de Biología animal del CSIC**

Documento:83575

Título:Estructura trófica de la macrofauna intermareal: Evolución en un área afectada por vertidos orgánicos.

Otro-Idioma-Título:Trophic structure of intertidal macrofauna: Evolution in an area affected by organic dumping.

Autores:Palacio, J. ; Mora, J. ; Lastra, M. ; Planas, M.

Lugar-Trabajo:Fac. Biol. Univ. Santiago de Compostela Dep. Biol. Anim., Santiago de Compostela (La Coruña), España ; Inst. Invest. Mar. [CSIC], Vigo (Pontevedra), España

Revista-Título:Publicaciones especiales. Instituto Español de Oceanografía

Datos-Fuente:1993, (11): 415-422, 18 Ref

ISSN:0214-7378

Idioma:ES

Tipo-Documento:AR

Modo-Documento:PO

Clasificación:251005 ; 330811

Descriptores:Contaminación de aguas marinas ; Contaminación industrial ; Fauna marina ; Indicadores biológicos ; Zona intermareal ; Ecología trófica Topónimos:Galicia ; España

---

<sup>14</sup>MOYA ANEGÓN, F. de. **Los sistemas integrados de gestión bibliotecaria**. Madrid: Anabad, 1995, p. 151.

**Base de datos de Electrotecnia, de ASINEL:**

**CONTADOR** 359  
**ACCN** 359  
**NUMPIE** 131054  
**NUMPIU** 87/30  
**TÍTULO** APLICACIÓN DE UN SISTEMA EXPERTO A LA QUÍMICA DE LOS SISTEMAS AGUA-VAPOR Y TRATAMIENTO DE AGUA DE UNA CENTRAL TÉRMICA  
**TITABR** SIS. EXPERTO QUÍMICA AGUA  
**TIPODO** PROYECTO PIE  
**EMPRES** UEF/RES  
**ENTCOL** I.I.T./COL  
**FEINIC** 080587  
**FEFINA** 070388  
**DURACI** 10  
**AUTORE** ZAPATA, JOSÉ LUIS/RES  
**SIGNA** BA/197  
**FEENT** 100788  
**ESTADO** TERMINADO  
**PRESUP** 14200  
**CLATEM** 2.3 EXPLOTACIÓN  
**DESTEM** SISTEMA EXPERTO > CENTRAL TÉRMICA > INVESTIGACIÓN > CONTROL QUÍMICO > FALLO  
**RESOBJ** EL OBJETIVO DE ESTE PROYECTO ES LA INVESTIGACIÓN, DESARROLLO, IMPLANTACIÓN Y PUESTA A PUNTO DE UN SISTEMA EXPERTO CAPAZ DE REALIZAR UN DIAGNÓSTICO Y PREDICCIÓN DEL COMPORTAMIENTO DE LOS PARÁMETROS QUE GOBIERNAN LA QUÍMICA DE LA CENTRAL DE MANERA QUE PUEDA PREDECIR ANOMALÍAS INCIPIENTES QU PUDIERAN DAR LUGAR A FALLO Y AL MISMO TIEMPO ASESORAR SOBRE LAS MEDIDAS A TOMAR PARA EVITARLO, O BIEN, SI HUBIESE OCURRIDO EL FALLO, IGULAMENTE ASISTIR AL PERSONAL ENCARGADO CORRESPONDIENTE SOBRE LAS ACCIONES A LLEVAR A CABO. 3. CONTENIDO Y MEDIOS DEL PROYECTO 3.1 DESCRIPCIÓN COMO SE HA DICHO ANTES, EL OBJETIVO DE ESTE PROYECTO SERÍA LA INVESTIGACIÓN, DESARROLLO, IMPLANTACIÓN Y PUESTA A PUNTO POR UNIÓN ELÉCTRICA FENOSA DE UN SISTEMA EXPERTO CAPAZ DE

REALIZAR UN DIAGNÓSTICO Y PREDICCIÓN DE COMPORTAMIENTO DE LOS PARÁMETROS QUE GOBIERNAN LA QUÍMICA DE LA CENTRAL DE MANERA QUE PUEDA PREDECIR ANOMALÍAS INCIPIENTES QUE PUDIERAN DAR LUGAR A FALLO Y AL MISMO TIEMPO ASESORAR SOBRE LAS MEDIDAS A TOMAR PARA EVITARLO, O BIEN, SI HUBIESE OCURRIDO EL FALLO, IGUALMENTE ASISTIR AL PERSONAL ENCARGADO CORRESPONDIENTE SOBRE LAS ACCIONES A LLEVAR A CABO. PARA LA CONSECUCIÓN DE ESTE OBJETIVO, EL PROYECTO CONSISTIRÍA EN UN DESARROLLO POR PASOS QUE ABARCARÍA, DESDE LA ADQUISICIÓN DE DATOS HASTA LA IMPLANTACIÓN FINAL DEL SISTEMA EXPERTO. ESTOS PASOS COMPONENTES DEL DESARROLLO DEL PROYECTO SERÍAN A GRANDES RASGOS LOS SIGUIENTES: a) RECOGIDA DE INFORMACIÓN EN UN MEDIO ADECUADO. ESTA TAREA CONSISTIRÍA EN RECOGER EN UN ORDENADOR LOS DATOS DE LOS PARÁMETROS QUÍMICOS QUE SE MONITORIZAN. ACTUALMENTE SE REGISTRAN EN PAPEL CONTINUO. EL TIPO DE ORDENADOR NECESARIO PARA ELLO DEPENDE, EVIDENTEMENTE DE LA CANTIDAD DE PARÁMETROS A REGISTRAR. NOSOTROS ESTIMAMOS QUE PARA LA CAPTURA DE INFORMACIÓN NECESARIA AL SISTEMA EXPERTO APLICADO A LA QUÍMICA DE LOS SISTEMAS AGUA-VAPOR Y TRATAMIENTO DE AGUA, BASTARÍAN ORDENADORES PERSONALES NIVEL IBM-AT. b) ESTUDIO DE LAS CARACTERÍSTICAS DE LOS PARÁMETROS. ANTES DE EMPEZAR CON CUALQUIER TIPO DE ESTUDIO ES CONVENIENTE FAMILIARIZARSE CON LOS DATOS, ÉSTO ES, VER SUS INCIDENCIAS, VALORES MÁXIMOS Y MÍNIMOS, MEDIAS, VARIANZAS, RECONOCIMIENTO DE VALORES CARACTERÍSTICOS Y COMO SON INFLUENCIADOS POR OTROS PARÁMETROS PARA ESTABLECER POSIBLES ESTADOS Y DEPENDENCIAS. ESTO ES UN ESTUDIO NADA SENCILLO QUE HAY QUE TENER MUY CLARO PARA LAS ETAPAS SUBSIGUIENTES DE DESARROLLO Y PARA SU REALIZACIÓN HAY QUE ECHAR MANO DE ESTUDIOS ESTADÍSTICOS, POSIBLE ANÁLISIS DE SERIES TEMPORALES,...., TODO ELLO EN BASE A DATOS DE COMPORTAMIENTO YA RECOGIDOS Y QUE EN LAS CENTRALES TÉRMICAS DE ANLLARES NO SERÍAN DIFÍCILES DE CONSEGUIR, PUES EXISTE UN ARCHIVO HISTÓRICO DE REGISTRADORES QUE FACILITARÍAN LA LABOR. c) DISPONER DE UN DISPOSIT

IVO INFORMÁTICO ADECUADO PARA REALIZAR EL ESTUDIO ANTERIOR. AQUÍ SÓLO SE PUEDE INDICAR QUE SE PODRÍA PENSAR EN UTILIZAR LOS MISMOS MEDIOS EMPLEADOS PARA LA TAREA 1. d) ESTUDIO Y DETERMINACIÓN DE LA ESTRUCTURA DE LA BASE DE CONOCIMIENTOS Y MOTOR DE INFERENCIAS DEL SISTEMA EXPERTO. CONSISTIRÍA ESTA TAREA UNA VEZ CONOCIDO EL COMPORTAMIENTO DE LOS DATOS EN REALIZAR UNA RECOPIACIÓN DEL CONOCIMIENTO QUE SE TIENE EN EL ÁREA, TANTO DEPENDENCIAS DETERMINISTAS COMO DE LÓGICA DIFUSA ENTRE LOS DIVERSOS PARÁMETROS Y CARACTERÍSTICAS DE LOS SISTEMAS FÍSICOS, ASÍ COMO UNA ESTRUCTURACIÓN DE DICHO CONOCIMIENTO DE TAL MANERA QUE UNA PERSONA QUE NO LO TUVIESE, PUDIESE ACTUAR COMO SI LO TUVIERA DE VERDAD. ÉSTO SERÍA, EN REALIDAD, EL NÚCLEO FUNDAMENTAL DEL DESARROLLO DEL SISTEMA EXPERTO Y SERÍA UNA ETAPA DE MUY ESTRECHA COLABORACIÓN CON EL PERSONAL TÉCNICO DE LA CENTRAL EN EL ÁREA CONCRETA, QUE ES QUIEN DE VERDAD TIENE LA EXPERIENCIA. e) DISPONER DE UN DISPOSITIVO INFORMÁTICO ADECUADO PARA LA IMPLANTACIÓN DE LA BASE DE CONOCIMIENTOS Y MOTOR DE INFERENCIAS DEL SISTEMA EXPERTO. ESTA TAREA CONSISTIRÍA EN IMPLANTAR INFORMÁTICAMENTE EL ESTUDIO REALIZADO EN LA TAREA ANTERIOR. APARTE DEL TRABAJO DE IMPLANTACIÓN, AQUÍ HABRÍA QUE DISPONER DE UN DISPOSITIVO INFORMÁTICO ADECUADO. ASÍ NOSOTROS ESTIMAMOS QUE HARÍA FALTA UN ORDENADOR PC NIVEL AT AL CUAL SE LE PODRÍA APLICAR ALGUNA HERRAMIENTA TAL COMO (PERSONAL CONSULTANT) PARA INSTALAR ESTE SISTEMA EXPERTO.

**Base de datos de Medio ambiente del MOPU**

Autor: International Clean Air Conference 1978 Brisbane, Australia

Título: Proceedings of the International Clean Air Conference : clean air, the continuing challenge : Brisbane, Australia, May 15-19, 1978 sponsored by the Clean Air Society of Australia and New Zealand ; edited by E. T. White, P. Hetherington, B. R. Thiele

Publicación: Ann Arbor, Michigan : Ann Arbor Science

Fecha: cop. 1978

Descripción-Física: XII, 780 p

ISBN: 0-250-40254-8

Otros-Autores-Editores: White, E. T. ed. / Hetherington, P. ed. / Thiele, B. R. ed. / Clean Air Society of Australia and New Zeland

Descriptores: Contaminación atmosférica / Control ambiental / Precipitadores electrostáticos / Tecnología / Contaminantes atmosféricos

Clasificación: B66 Control de la contaminación atmosférica

Signatura: D-2795

**Base de de datos de Mujer, del Instituto de la Mujer:**

SIG .0LI001790  
AUT .BASUALDO, ANA  
TIT .LA MUJER EN LA IMAGEN PUBLICITARIA  
FUE .CUENTA Y RAZÓN; 50  
AÑO .1989  
COL .PP. 21-24  
DES .IMAGEN DE LA MUJER/TELEVISIÓN/PUBLICIDAD  
RES .ESTE ESTUDIO INTENTA ANALIZAR EL MODO EN QUE APARECE LA  
.MUJER ACTUAL EN LOS ANUNCIOS TELEVISIVOS. UNA PRIMERA  
.APROXIMACIÓN NOS DESCUBRE QUE NO HAY NINGÚN ANUNCIO EN  
.QUE LA VOZ QUE RELATA "EN OFF" NO SEA MASCULINA. ESTO  
.SUPONE YA UNA PRIMERA E IMPORTANTE DIFERENCIACIÓN DE  
.ROLES SEXUALES, DONDE LA VOZ QUE PRESENTA EL PRODUCTO Y  
.LO VENDE ES MASCULINA, MIENTRAS QUE LAS MUJERES SE  
.MANTIENEN EN UN SEGUNDO PLANO "PEGADAS" A LOS OBJETOS  
.QUE YA HAN SIDO CREADOS SI PROFUNDIZAMOS MÁS EN EL  
.ANÁLISIS DE LAS DISTINTAS IMÁGENES DE LA MUJER EN LOS  
.ANUNCIOS VEMOS QUE AÚN ES MAYORITARIA LA IMAGEN TRADI-  
.CIONAL DE LA MUJER COMO AMA DE CASA QUE TIENE POR MISIÓN  
.TENER TODO DISPUESTO EN EL HOGAR Y CUIDAR DE LOS HIJOS,  
.EXISTE OTRA IMAGEN (YA CLÁSICA) QUE PODRÍAMOS DENOMINAR  
.COMO DE "OBJETO DECORATIVO" QUE ACOMPAÑA AL PRODUCTO  
.A VENDER. SIN EMBARGO, ÚLTIMAMENTE SE ESTÁ DESARRO-  
.LLANDO UN NUEVO TIPO DE "MUJER PUBLICITARIA" QUE, FRE-  
.CUENTEMENTE APARECE LIGADA A LOS MUNDOS DE LA TECNO-  
.LOGÍA Y DE LA ECONOMÍA, Y QUE ES AQUELLA QUE NO SE  
.DISTINGUE NI POR EL PAPEL QUE JUEGA NI POR SU CAPACIDAD DE  
.LOS HOMBRES QUE TAMBIÉN APARECEN EN ESOS ANUNCIOS, NO  
.OBSTANTE, INCLUSO EN ESTOS CASOS LA VOZ QUE VENDE (LA VOZ  
.EN OFF) SIGUE SIENDO MASCULINA  
LUG .GENÉRICO

**Base de datos de Psicología, del CSIC**

Núm. registro.....: 232102

Autores.....: BARRIGA, SILVERIO

Título del artículo.: ROL DEL PSICOLOGO EN LOS SERVICIOS SOCIALES

Lugar de trabajo....: UNIV. SEVILLA, DEP. PSICOLOGIA SOCIAL, ESPAÑA

ISSN de la revista...: 1131-6225

Título Revista.....: REVISTA DE PSICOLOGIA SOCIAL APLICADA

Datos fuente.....: 1991, 1 (2-3): 101-112, 42 REF

Tipo documento.....: AR

Modo del documento..: PO

Lengua.....: ES

Clasificación.....: 651306

Localización.....: ISOC

Congreso.....: JORNADAS DE ALMUÑECAR. 1991

Notas.....: RESUMEN (ES, EN)

Descriptores.....: ROLES;PSICOLOGOS;SERVICIOS SOCIALES;BIENESTAR SOCIAL;ENSEÑANZA PROFESIONAL

Resumen.....: EL PRESENTE ARTICULO EL AUTOR REFLEXIONA SOBRE EL ROL DEL PSICOLOGO EN LOS SERVICIOS SOCIALES TANTO EN LOS ESPECIALIZADOS COMO EN LOS COMUNITARIOS, PARA PERFILAR LOS PRINCIPALES ASPECTOS DEL MENCIONADO ROL PARTE DE LA CONSIDERACION DE LOS SERVICIOS SOCIALES EN EL MARCO DEL BIENESTAR SOCIAL.-- LA CONSIDERACION DE LOS COMPONENTES DEL ROL DEL PSICOLOGO EN ESTOS AMBITOS ES LA BASE PARA SEÑALAR LOS ASPECTOS BASICOS QUE DEBE CONTEMPLAR LA FORMACION DE LOS PSICOLOGOS QUE VAYAN A EJERCER SU PRÁCTICA PROFESIONAL EN ESTE CAMPO. (A.)

**Base de datos de Servicios sociales**

REF1 :014048

FECHENT :930922

TIPODOC :AR

AUTORPER:HORWATH, Caroline C.

TIT :NUTRITION GOALS FOR OLDER ADULTS : A REVIEW

TITULO :Nutrition goals for older adults : a review / Caroline C. Horwath

LUGAR :Washington, D.C.

EDITOR :Gerontological Society of America

FECHA :1991

AÑO :1991

COLACION:< 11 > p. ; 28 cm

NOTBIB :Bibliografía: p. 819-821

FUENTE :THE GERONTOLOGIST, VOL 31 N 6, DEC 1991. -- PP. 811-821

LENGUA :eng

PAIS :USA

ISSN :0016-9013

DESMAT :TERCERA EDAD/NUTRICION/DIETETICA/ENFERMEDADES DEL APARATO CIRCULATORIO/ENFERMEDADES ENDOCRINAS/CANCER/PREVEN-  
CION/EDUCACION SANITARIA

ENT-SEC :GERONTOLOGICAL SOCIETY OF AMERICA, ed.

CLASIF :R/Z

#### **4.4. Búsqueda documental**

El proceso de búsqueda o recuperación de información conlleva una serie de etapas interrelacionadas entre sí y que siguen este orden:

1. Análisis de los conceptos de la consulta
2. Elección del tipo de lenguaje a utilizar
3. Definición de la estrategia de búsqueda
4. Establecimiento de enlaces sintácticos entre los términos de búsqueda
5. Visualización y evaluación de los resultados
6. Exportación de los registros

##### **4.4.1. Análisis de los conceptos de la consulta**

Fase en la que se establece la comparación entre la información almacenada y las demandas de información. Su nivel de coincidencia permitirá la recuperación de una información más o menos precisa, en función de la correspondencia entre la pregunta formulada y los términos utilizados para indizar y almacenar el documento.

Se seleccionan los términos presentes en la pregunta del usuario y se establece una lista de posibles términos que pueden usarse para describir la materia, comprobando en índices actualizados si son o no apropiados, o bien realizando la comprobación con un tesoro si hay alguno disponible sobre el tema<sup>15</sup>. En una fase posterior, los conceptos de la lista se reunirán en una ecuación de búsqueda. El procedimiento para realizar esta

---

<sup>15</sup>El documentalista profundizará en el tema de la pregunta consultando repertorios, diccionarios, revistas especializadas, etc. o bien un tesoro, si la base dispone de uno especializado. La capacidad de inducción de este lenguaje documental facilita dicha profundización.

tarea es similar al que realiza el documentalista en el momento de indizar los documentos para alimentar la base de datos, pero existen matices diferenciadores, como aclara Van Slype<sup>16</sup>, *... la enumeración de los conceptos que componen una pregunta exige una aproximación muy diferente de la que se lleva a cabo cuando se hace la lista de los conceptos que caracterizan el contenido de un documento. La indización de un documento es una operación reductora: el contenido de un documento de cinco a diez páginas queda considerablemente condensado cuando se representa por un número de ocho a doce descriptores, conforme a una tasa del 1/100 al 1/1000. Por el contrario, lo más frecuente es que la formulación de una pregunta produzca una extensión del enunciado final: una petición expresada por el enunciado de un número de dos a cinco conceptos, se enriquece de forma que se obtienen, como media, de cinco a doce palabras clave y/o descriptores, de acuerdo con una tasa de expansión del 100 al 300%.*

El incremento en el número de conceptos posibles a considerar en el proceso de recuperación tiene su razón de ser en el hecho de que un documento puede responder a una consulta considerando el tema de diversas formas. Por ejemplo, el concepto *azúcar*, puede ser abordado desde los puntos de vista de la *dietética*, la *producción de caña de azúcar*, su *industrialización*, etc.

#### **4.4.2. Elección del tipo de lenguaje a utilizar**

Es posible elegir entre el empleo de lenguaje controlado o de lenguaje libre. En el primer caso se traducen los conceptos de la pregunta a descriptores o encabezamientos de materia, si elegimos como tal lenguaje un tesaurus o una lista de encabezamientos. En

---

<sup>16</sup>SLYPE, G.v. Op. cit. p. 162.

el segundo caso se mantienen los términos de la pregunta, sin normalizar. Si se trata de documentos extensos o si las consultas son algo complejas, la recuperación en lenguaje libre puede resultar muy costosa en cuanto a tiempo de respuesta.

#### **4.4.3. Definición de la estrategia de búsqueda**

Definición de la ecuación de búsqueda de acuerdo con la filosofía de la base de datos objeto de la consulta y atendiendo a la sintaxis propia del Servicio de Recuperación de Información de dicha base.

Hartley define la estrategia como *conjunto de decisiones y acciones tomadas para realizar una búsqueda*<sup>17</sup>, y sostiene que dichas decisiones afectan el resultado conseguido.

Lo habitual es plantear una estrategia inicial que constituya un primer intento de diálogo con el sistema y, en función de los resultados obtenidos, replantear la ecuación para localizar otros términos más genéricos, más específicos o relacionados, que puedan sugerir otros conceptos útiles para la interrogación.

Ajeno o no a la experiencia del usuario o del documentalista en la elaboración de estrategias de búsqueda, es conveniente comenzar la consulta con un planteamiento de la cuestión sencillo y, una vez visualizados los resultados de la ecuación inicial, se puede ir complicando hasta obtener la información pertinente.

---

<sup>17</sup>HARTLEY, R.J. *Online searching: principles & practice*. London [etc.]: Bowker-Saur, 1990, p. 153.

En la preparación de la estrategia de búsqueda se sigue el siguiente proceso:

- Recopilación o listado de los términos facilitados por el usuario.
- Búsqueda de sinónimos o cuasisinónimos en diccionarios especializados, revistas especializadas, etc. Si en la etapa anterior se decidió seguir el procedimiento en lenguaje controlado, se pueden ampliar los conceptos dados por el usuario y comprobar si los términos seleccionados están en el tesoro. Si se elige la búsqueda en lenguaje libre, el proceso se reduce a la localización de sinónimos mencionada, para tratar de tener en cuenta cualquier término relacionado con el tema de la consulta.

Dependiendo de su complejidad, la formulación de la estrategia de búsqueda lleva sólo unos minutos, media hora e incluso más tiempo si la pregunta es compleja y requiere muchas sentencias. Algunas consultas son para una información precisa, otras son más generales, e incluso puede ser necesario hacer un barrido o ramoneo, *browsing* en inglés, a través de la información almacenada en la base, para cubrir determinadas necesidades insuficientemente perfiladas en la consulta. Foskett sostiene que este tipo de consulta se lleva a cabo también cuando el lector no tiene una necesidad particular en mente y trata de encontrar documentos por azar y considera de gran utilidad esta posibilidad que ofrecen algunos sistemas de recuperación de información, dado que muchos de los más significativos descubrimientos científicos se han realizado como resultado de lo que denomina *serendipity*, que define como *la facultad de hacer felices e inesperados descubrimientos por accidente*<sup>18</sup>.

La práctica del *browsing* se viene realizando en los procesos de recuperación de información desde que se generalizó el uso del CD-ROM como soporte de almacenamien-

---

<sup>18</sup>FOSKETT, A.C. Op. cit. p. 26.

to de la información, que abarató el uso de los sistemas de recuperación y que permite utilizar las bases de datos sin presiones del factor tiempo características de la consulta en bases de datos en línea.

#### **4.4.4. Establecimiento de enlaces sintácticos entre los términos de búsqueda.**

*En el proceso de recuperación de información no es habitual realizar una búsqueda en la que se utilice un concepto simple, ya que este tipo de cuestiones se resuelven sin necesidad de preparar estrategia alguna, simplemente consultando el índice o el resumen de un documento. Sin embargo, por lo general, para una búsqueda se requieren entre dos y tres conceptos, que será necesario combinar mediante algún procedimiento sintáctico.*

##### **4.4.4.1. Álgebra de Boole**

Entre los métodos de búsqueda, figura el álgebra de Boole, que es el elegido por la mayoría de las bases de datos cuando se quieren realizar búsquedas a partir de varias palabras. Se trata de un método de lógica simbólica basado en el empleo de conectores u operadores lógicos, ideados por Boole.

Boole, matemático británico del siglo XIX, (1818-1864), empleó símbolos matemáticos para expresar procesos lógicos. Los operadores de unión [o], intersección [y] negación [no], proveen una flexible forma de combinar dos o más términos o conceptos para conseguir en base a ellos el concepto más representativo de determinadas consultas. Su obra fue desarrollada por Venn (1834-1923) quien representó las relaciones

lógicas del álgebra booleana mediante diagramas adoptando los círculos de Euler (1707-1783), matemático suizo que introdujo la técnica de expresar relaciones lógicas gráficamente un siglo antes de que Boole inventase sus operadores lógicos.

La conexión de los términos presentes en una pregunta se realiza de la siguiente forma con los operadores lógicos:

### *Operador de intersección*

Permite la intersección entre dos o más conceptos para poder recuperar aquellos documentos en los que estén presentes todos ellos simultáneamente. Sirve, por tanto, para reducir el número de items resultantes de una búsqueda. Ej.:

*costas rocosas y costas arenosas*



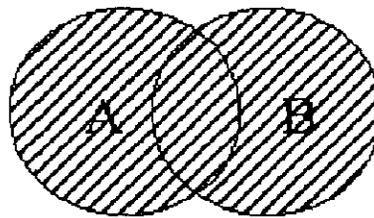
El sistema recuperará en este caso sólo los documentos en los que aparezcan los términos costas rocosas y costas arenosas.

### *Operador de unión*

Agrupar palabras clave que, o bien son sinónimas, o describen un mismo concepto.

Ej.:

*costas rocosas o costas arenosas*



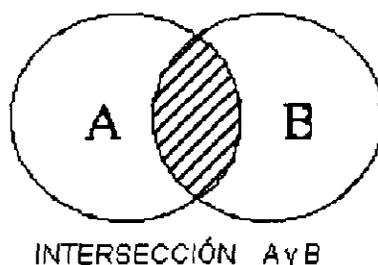
UNIÓN O INCLUSIÓN  $A \cup B$

El operador de unión sirve para ampliar la búsqueda, de tal manera que, en el ejemplo, el sistema recuperará los documentos referidos tanto a costas rocosas como a costas arenosas.

### *Operador de exclusión*

El operador **no** se usa para excluir un concepto en particular, o un subconjunto dentro de un conjunto más amplio. Ej.:

*costas arenosas no costas rocosas*



En este caso, se trata de una la búsqueda sobre costas arenosas, y se excluyen las costas que no sean playas. Este operador debe utilizarse con precaución porque se puede perder información relevante. En el ejemplo, al excluir documentos en donde aparezcan costas rocosas, pueden excluirse igualmente documentos que traten otro tipo de costas, incluidas las arenosas, que podrían ser de interés para el usuario.

#### 4.4.4.2. Truncamiento

El método de búsqueda por truncamiento consiste en la formulación de la estrategia empleando fragmentos de palabras precedidos o sucedidos de un signo que permite al sistema reconocer y tomar como pertinente cualquier término que contenga letras situadas antes o después de dicho signo. *El truncamiento* -escribe Van Slype- *evita al usuario tener que introducir manualmente dentro de la ecuación todas las flexiones gramaticales, y a veces todas las formas derivadas de una misma raíz*<sup>19</sup>.

Los símbolos más utilizados para este recurso son el asterisco [\*], la interrogación [?] y el signo de interjección [!] que el ordenador toma como sustitutos de cualquier carácter.

---

<sup>19</sup>SLYPE, G.v. Op. cit., p. 171.

Existen tres tipos de truncamiento: de sufijo, de prefijo y de infijo. El truncamiento de sufijo permite recuperar bloques que contienen palabras que empiezan o presentan coincidencia con la petición. Por ejemplo, *graf\**, abarcaría los términos:

graf-ía  
gráf-ico  
graf-ología

El truncamiento de prefijo, *\*grafía*, localizaría:

grafía  
eco-grafía  
homo-grafía

El de infijo *\*graf\**, consideraría cualquiera de las posibilidades señaladas.

El truncamiento es un método de búsqueda utilizado en bases de datos con indización en lenguaje libre con objeto de aumentar la exhaustividad, aunque con este recurso se corre el riesgo de recuperar documentos irrelevantes. Por ejemplo, en la búsqueda que realizamos para esta tesis en la base de datos del Centro de Documentación de la Mujer, solicitamos al sistema documentos relativos a *deporte*, truncando la raíz *deport\**, y entre las referencias analizadas encontramos una que se refería a mujeres *deportadas*, palabra morfológicamente próxima, pero sin ninguna relación semántica con el tema de consulta.

#### **4.4.4.3. Enlaces de proximidad**

El método de búsqueda por enlaces de proximidad permite localizar los términos de la consulta dentro del texto siempre que aparezcan en un lugar predeterminado del mismo. Se caracteriza por limitar el número máximo de palabras que pueden existir entre las especificadas en una determinada petición.

Existe gran variedad de posibilidades de establecer enlaces de proximidad. Entre los más comunes:

- Recuperación tan sólo de aquellos documentos en los que los términos de la pregunta aparezcan juntos, uno al lado de otro, sin ninguna palabra entre ellos.
- Introducción de un operador que considere la posibilidad de recuperar términos separados por un número determinado de palabras.
- Recuperación solamente de aquellos documentos en los que los términos que componen la pregunta aparezcan dentro de una misma frase.
- Operador para seleccionar términos separados entre sí por un número determinado de términos en un orden preestablecido.
- Operador para seleccionar términos separados entre sí por un número determinado de términos en cualquier orden.

#### 4.4.4.4. Método de ponderación o recuperación por pesos

La indización implica habitualmente una decisión dicotómica. Un término es asignado a un documento o no. En vez de tomar tal decisión es posible introducir valoraciones cuantitativas, es decir, *darle un valor a los términos de indización para indicar su importancia relativa para representar la materia del documento*<sup>20</sup>. Este recurso, conocido como indización por pesos y también como método de ponderación, *neutraliza los efectos de la combinación binaria de términos que realiza el álgebra de Boole, con cuyo procedimiento un documento dado o es pertinente o no es pertinente, sin posibles matices*<sup>21</sup>.

Es lógico suponer que entre los términos elegidos para indizar un documento dado, así como los que lo representan en el momento de la recuperación, unos son más importantes que otros. Durante años los investigadores han tratado de conseguir métodos para asignar un valor a estos descriptores con objeto de mejorar los resultados al recuperar la información. Luhn fue pionero -como en tantos otros avances tecnológicos- en el desarrollo de algoritmos que permitían la asignación de un valor a los términos de indización. Moya Anegón escribe: *Como conclusión a las aportaciones de Luhn se puede decir que existe una gradación posiblemente cuantificable de la importancia que cada término de indización tiene en la representación del contenido de los documentos*<sup>22</sup>.

---

<sup>20</sup>LANCASTER, F.W. *Vocabulary control for information retrieval*. Washington, D.C.: Information Resources Press, 1972, p. 132.

<sup>21</sup>SLYPE, G.v. *Op. cit.*, pp. 174.

<sup>22</sup>MOYA ANEGÓN, F. de. *Op. cit.*, p. 168.

Salton y McGill<sup>23</sup> formularon tres algoritmos que el programa calcula en el momento de la recuperación. En todos ellos el sistema extrae un conjunto de referencias ordenadas en función de la carga semántica que el propio sistema atribuye a cada una de dichas referencias, a partir de la pregunta del usuario. También fue Salton quien, con sus colegas de la Universidad de Cornell, llevó a cabo el proyecto SMART, la primera investigación realizada sobre el método de ponderación.

Si se procede a recuperar mediante pesos, pueden darse dos posibilidades: que los documentos se hayan indizado también por pesos o que no se haya realizado su ponderación en esta fase del tratamiento documental. En el primer caso, se establece la coincidencia entre el valor de indización con el deseado para la recuperación. Si se solicitan documentos en los cuales un término tenga el peso 3, el sistema responderá, exclusivamente, con aquellos documentos cuyos descriptores respondan a dicho valor. En el caso de que no se haya realizado la ponderación en la fase de indización, se puede aplicar en el momento de la recuperación, en función directa de la frecuencia de aparición de los términos solicitados, en cada documento o en la colección.

La búsqueda por pesos aporta flexibilidad a la operación porque ofrece la posibilidad de variar las estrategias de acuerdo con las necesidades de precisión o de exhaustividad. Cuando deseamos un alto índice de precisión lo único que tenemos que hacer es pedir al sistema que se limite a darnos sólo las referencias de aquellos documentos cuyos descriptores tengan un alto nivel de ponderación. Cuando lo que necesitamos es un nivel de exhaustividad alto, se ignoran los valores dados a los descriptores en la fase de indización, de tal forma que recuperamos la información de cualquier *peso* sobre la materia deseada. En opinión de Lancaster, *con este recurso se*

---

<sup>23</sup>SALTON, G. & MCGILL, M.J. Op. cit., pp. 350-355.

*evita la recuperación de documentos que sólo hacen mención de la materia requerida de forma marginal<sup>24</sup>.*

Maron, M.G y KUHNS, J.L.<sup>25</sup> proponen la siguiente valoración de los descriptores:

<b>Peso</b>	<b>Descripción</b>	<b>Utilización</b>
8/8	Materia principal	El término es muy específico y cubre por completo la materia principal del documento
7/8	Materia principal	El término es específico y cubre casi por completo la materia principal del documento
6/8	Materia más genérica	El término es demasiado amplio y cubre una materia principal
5/8	Otros términos importantes	Términos que podrían utilizarse en indización binaria pero que no representan a la materia principal
4/8	Materia menos genérica	El término se refiere al tema tratado en el documento pero es demasiado específico para representar una materia principal
3/8	Materia secundaria	Incluye información sobre resultados de experimentos, métodos, usos, etc.
2/8	Otras materias	Otros identificadores
1/8	Apenas relevante	Materias que el indizador no considera relevantes pero que extrae porque cree que algún usuario puede considerarlas pertinentes

---

<sup>24</sup>LANCASTER, F.W. Op. cit., p. 132.

<sup>25</sup>MARON, M.E. & KUHNS, J.L. On relevance, probabilistic indexing and information retrieval. **Journal of the ACM**, v. 7, nº 3 (1960) pp. 216-244.

Si la búsqueda por pesos es ya de por sí considerada subjetiva, la valoración semántica que proponen Maron y Kuhns puede resultar demasiado minuciosa para su aplicación práctica. Establecer la distinción entre valores tan aproximados como 7 y 8, por ejemplo, resultaría tarea difícilmente realizable en un caso concreto.

Se podría considerar una alternativa en la que se atribuyeran a los descriptores pesos entre 3 y 1, con los cuales los documentos recibirían una valoración de acuerdo con la profundidad con que tratan una materia en particular y darían opción al usuario de limitar la búsqueda a los documentos pertinentes. El peso 3 podría utilizarse cuando un descriptor representa con exactitud el contenido del documento; el valor 2 cuando el concepto es importante, pero no representa exactamente la información esencial del documento; el valor 1, por último, se utilizaría para medir un concepto que se acepta como término de indización o de recuperación, aunque no representa el tema central del documento.

#### **4.4.4.5. Método estadístico**

Consiste en la asignación de términos a un documento en base a la frecuencia de aparición de dichos términos en el documento y en la colección documental a la que pertenece.

En los programas de recuperación automática por métodos estadísticos, todas las palabras que componen el texto son objeto de análisis, a excepción de las palabras vacías, que se recogen en una lista ("stop word list" o lista de palabras vacías), y que el ordenador va rechazando a medida que lee secuencialmente. La lista de palabras vacías

se compone de conjunciones, adverbios, artículos, adjetivos, pronombres personales, relativos, posesivos, demostrativos e indefinidos. Las preposiciones, además de su papel conectivo o relacionante, pueden tener valor sustantivo, por lo que no deben considerarse palabras vacías sin previo análisis del contexto.

El método estadístico tiene su fundamento en la teoría de la probabilidad, y parte de dos hipótesis: a) que los términos que aparecen con mucha frecuencia en un documento son representativos de su contenido; b) que aquellos términos que tienen un bajo nivel de frecuencia en un conjunto documental no son útiles para la recuperación.

Salton y McGill<sup>26</sup> atribuyen a Luhn el origen de este procedimiento de recuperación, y añaden al análisis estadístico lo que denominan valor de discriminación que, relativiza la valoración de los términos en función de su frecuencia de aparición en los documentos o en las bases de datos. Salton y McGill sostienen que las palabras que aparecen pocas veces en una base de datos tienen un valor discriminativo más alto que aquellas que son compartidas por muchos registros. En la base de datos del Centro de documentación de la Mujer, por ejemplo, la palabra *mujer* es tan frecuente en los documentos que, aunque tiene un gran valor semántico en dicho centro, carece de valor discriminativo por su altísimo nivel de frecuencia.

#### **4.4.4.6. Método de recuperación mediante tesoro**

Como explicamos en el capítulo segundo de esta tesis doctoral, un tesoro es una lista de términos controlada que se utiliza para indizar los documentos, así como para su

---

<sup>26</sup>SALTON, G. & McGill, M.J. Op. cit., p. 66.

recuperación. La estructura jerárquica en la que organiza los descriptores favorece un despliegue de relaciones semánticas -no sólo morfológicas, como ocurre con la recuperación por términos- que le otorgan su capacidad de inducción.

Algunas bases de datos permiten el acceso al tesoro a través del interfaz; en aquellas en las que no existe este recurso se hace necesaria su consulta en papel. De las seis bases en las que hemos realizado el test, objeto de esta investigación, solamente una, la de *Asuntos sociales*, carece de tal recurso.

En los centros que tienen incorporado el tesoro existe un comando que da acceso al mismo; el siguiente paso consiste en visualizar los descriptores como si se tratara de un índice, o teclear el descriptor que se estime adecuado para recuperar determinada información, el cual aparece en pantalla rodeado de sus términos más genéricos y más específicos, así como de aquellos que, por asociación o sinonimia, estén relacionados con él. Entre todos ellos se elige la opción más adecuada al perfil de búsqueda. Si buscamos, por ejemplo, por *comunicación*, podemos encontrar el término específico *comunicación no verbal*, e incluso el más específico aún *expresión corporal*; al propio tiempo el tesoro induce a búsquedas asociadas al presentar junto a tales descriptores otros términos relacionados con ellos por asociación de ideas, como, por ejemplo, *mensaje*.

Cuando no es posible el acceso al tesoro mediante el interfaz, se seleccionan manualmente el o los términos susceptibles de recuperar información pertinente, y se teclean en la pantalla en la que se esté preparando la ecuación de búsqueda.

Tanto si se puede consultar el tesoro en línea como si ha de hacerse manualmente, es posible reducir o ampliar la búsqueda mediante el uso de descriptores más o menos

específicos así como combinando el empleo de los operadores booleanos y otros recursos mencionados en este capítulo.

#### **4.4.5. Visualización y evaluación de los resultados**

En los SRI en línea, esto es, aquellos que permiten hacer consultas y obtener respuestas en tiempo real, la visualización de los resultados puede hacerse al concluir el proceso de búsqueda o cuando se consiguen los primeros registros. Es habitual la visualización en esta segunda fase porque permite mejorar la formulación de la ecuación y optimizar los resultados definitivos.

La evaluación de los resultados se puede hacer en función de la precisión y de la exhaustividad, parámetros que analizaremos en el capítulo sexto de este trabajo. Nos limitamos aquí a citar los cuatro casos que se pueden dar al evaluar las respuestas:

1. Que las referencias sean pertinentes y su número adecuado al perfil de búsqueda.
2. Que las referencias sean pertinentes, pero demasiadas para determinado perfil.
3. Que, junto con las referencias pertinentes, se recupere información no pertinente.
4. Que se recuperen referencias pertinentes pero insuficientes para satisfacer una consulta concreta.

Salvo en el primer caso, en todos los demás sería necesario utilizar algún recurso que mejorara la pertinencia. En el segundo, se puede limitar el número de referencias restringiendo la búsqueda a aquellos documentos que se limiten a un período de tiempo

determinado, a un tipo de documento en particular o a un idioma. Cuando se formula una ecuación que genera ruido, como ocurre en el tercer caso, se puede restringir el área de búsqueda usando descriptores más específicos que los que se seleccionaron en la estrategia original. Por último, en el cuarto supuesto, se da el caso contrario. El silencio documental se puede neutralizar incluyendo en la estrategia términos genéricos y asociados a los formulados en la primera ecuación.

En SRI que practiquen la recuperación por pesos se pueden controlar los resultados de las dos últimas posibilidades expuestas. Cuando el número de documentos es demasiado alto, se procede a la ponderación de los descriptores; por el contrario, cuando su número es insuficiente, se omite tal recurso, consiguiendo así un generoso aumento de referencias.

#### **4.4.6. Exportación de los registros**

La exportación consiste en el envío de los resultados obtenidos de la búsqueda desde la base de datos a un soporte, para su posterior utilización. Dicha exportación puede hacerse mediante la impresión de las referencias en papel o grabándolas en un fichero. No todas las bases de datos ofrecen ambas opciones al usuario; en algunas cabe sólo la posibilidad de grabar los registros a un fichero. Si se da el caso de que la masa de documentos recuperados es densa y que, además, se ha solicitado la exportación de los registros con todos sus campos (incluido el de resumen), la labor de analizar su pertinencia resulta sensiblemente más incómoda que si se trabaja con los registros impresos.

Aunque en la fase de visualización y evaluación de los resultados se valora la información obtenida, sólo cuando se exportan los resultados se puede proceder a un análisis detenido de las referencias y estimar en qué medida se han cumplido la predicciones del consultante sobre la base de datos, con el contenido de la misma. En base a esas expectativas el usuario considerará o no pertinente la documentación recuperada.

## Capítulo 5

Análisis semántico y estructural de diferentes tesauros y características de las bases de datos en las que se utilizan

## 5.1. Introducción

El planteamiento teórico de las distintas técnicas de recuperación de información que hemos realizado en el capítulo anterior, abre un abanico de posibilidades de elección para proceder a las búsquedas que habremos de desarrollar en el test, objeto de esta tesis doctoral, y que presentamos en el sexto capítulo. Igualmente, para poder llevar a cabo dicha experimentación, procedemos en el presente apartado a hacer un análisis semántico y estructural de los seis tesauros utilizados para dicho test, y a examinar las características de las seis bases de datos consultadas.

Los tesauros y las bases de datos que componen la muestra son:

- *Tesauro ICYT de Biología animal*<sup>1</sup>, de uso en la base de datos sobre la materia del CSIC.

- *Tesauro Electrotécnico*<sup>2</sup>, útil al servicio de la base de datos PIE (Programa de Investigación y Desarrollo Electrotécnico).

- *Tesauro de Medio ambiente*<sup>3</sup>, lenguaje de indización y recuperación de la Base de datos Biblioma, del Centro de documentación del Ministerio de Medio ambiente.

---

<sup>1</sup>REY ROCHA, J. *Tesauro ICYT de Biología animal*. Madrid: CINDOC, 1995.

<sup>2</sup>TESAURO *Electrotécnico*. Madrid: Asociación de Investigación Industrial Eléctrica (ASINEL), 1993.

<sup>3</sup>TESAURO *de Medio Ambiente*. 2ª ed. Madrid: Ministerio de Obras Públicas y Urbanismo, 1990.

- *Tesaurus "Mujer"*<sup>4</sup>, de empleo en el Centro de documentación de la Mujer, dependiente del Ministerio de Asuntos sociales.

- *Tesaurus ISOC de Psicología*<sup>5</sup>, construido para tratar la información que sobre la materia tiene la base de datos del CSIC.

- *Tesaurus de Asuntos sociales*<sup>6</sup>, utilizado en la Biblioteca del Instituto de Migraciones y Servicios sociales.

## 5.2. Criterios de evaluación

Para llevar a cabo la valoración de los tesauros seguimos los criterios de la *Norma UNE 50-106-90*, equivalente a *ISO-2788-1986 Directrices para el establecimiento y desarrollo de tesauros monolingües*<sup>7</sup>, así como los indicados por algunos autores, entre los que destacamos a Lancaster<sup>8</sup>. Los parámetros que se van a medir son los siguientes:

- **Composición:** Se refiere a los índices que forman el tesaurus y a la existencia o no de una introducción con información sobre sus características y ámbito de aplicación.

---

<sup>4</sup>TESAURO "Mujer". Madrid: Instituto de la Mujer, 1993.

<sup>5</sup>TESAURO ISOC de Psicología. Madrid: CINDOC, 1995.

<sup>6</sup>TESAURO de Servicios sociales. 2ª ed. Madrid: Ministerio de Asuntos sociales, 1994.

<sup>7</sup>UNE-50-106-90. Directrices para el establecimiento y desarrollo de tesauros monolingües. AENOR, 1990.

<sup>8</sup>LANCASTER, F.W. Evaluación de tesauros. En: *El control del vocabulario en la recuperación de información*. Valencia: Universitat, 1995, pp. 173-174.

● **Tamaño:** Número de descriptores y de no descriptores que componen la terminología. Los no descriptores no se asignan a los documentos en el momento de indizar ni se pueden, por lo tanto, utilizar para recuperar información, pero sirven de referencia para el usuario y van precedidos de la sigla USE, que remite al término descriptor.

● **Relaciones de equivalencia.** El valor deseable fluctúa según las fuentes, *algunas la sitúan entre un 0'5 y 2, mientras que otras señalan simplemente que el valor mínimo debe ser mayor a 1<sup>º</sup>.*

● **Relaciones jerárquicas y asociativas,** o razón de enriquecimiento que denominan algunos autores. Mide cuantitativamente la proporción existente entre el número de relaciones jerárquicas y/o asociativas y el número total de descriptores. La media aconsejada oscila entre 2 y 5 relaciones por descriptor.

● **Reciprocidad de las relaciones semánticas.** Se valora en este punto si se ha tenido en cuenta la necesaria reciprocidad de las relaciones de equivalencia, jerárquicas y asociativas en el momento de la construcción del tesoro.

● **Notas aclaratorias, definatorias o de aplicación, también denominadas notas de alcance.** Se utilizan para aclarar términos que tienen varios significados en lenguaje natural o cuando un descriptor se utiliza con un sentido especial en el tesoro. Las notas que presentan una definición completa del descriptor se aplican a términos técnicos o especializados, característicos de una jerga o de un contexto social o institucional.

---

<sup>9</sup>ALVARO BERMEJO, C.; VILLAGRÁ RUBIO, A.; SORLI ROJO, A. Desarrollo de lenguajes formalizados en lengua española: II. Evaluación de los tesauros disponibles en lengua española. *Revista Española de Documentación Científica*, v. 12, nº 3, p. 293.

Las citadas notas constituyen un valor cuantificable numéricamente; cuantas menos notas incluya un tesoro, mayor será el número de descriptores que puedan generar ambigüedad. Si en el lenguaje incluyen notas menos del 10% de los descriptores, se puede pensar que hay ambigüedad notable.

- **Número de niveles jerárquicos.** Permiten conocer el nivel de especificidad de los términos que componen el tesoro. Se trata de un valor relativo en función de la especificidad de la materia.

- **Morfología de las palabras.** Se refiere al uso que se hace en el tesoro de las formas de las palabras; del empleo del singular y el plural y de si los descriptores se expresan por medio de entradas directas, sin inversión de términos, respetando el orden natural de las palabras.

- **Aspectos estéticos de composición y tipografía.** Se tendrán en cuenta en este punto las cualidades formales del tesoro, tales como tipo de letra utilizado para representar los términos, sangrados, etc. La norma UNE-50-106-90<sup>10</sup>, hace mención expresa de la forma en la que deben destacarse los descriptores de los no descriptores:

*a) Los términos preferentes se imprimen en letras mayúsculas, por ejemplo:*

*COCHES  
ANIMALES*

---

<sup>10</sup>UNE 50-106-90. Directrices para el establecimiento y desarrollo de tesauros monolingües. Madrid: AENOR, 1990, pp. 467-468.

b) Los términos no preferentes se imprimen en minúsculas, salvo en los casos de nombres propios que requieran una inicial en mayúscula o las abreviaturas o acrónimos, que siempre deben escribirse en mayúsculas, por ejemplo:

*COCHES*

*UP automóviles*

*ANIMALES*

*UP fauna*

*ORGANIZACIÓN MUNDIAL DE A SALUD*

*UP OMS*

● **Nivel de precoordinación** o número medio de términos significativos por descriptor. Esta tasa ( $t$ ), cuyos valores aconsejados oscilan entre 1'5 y 2, se obtiene a través del recuento de la frecuencia de los descriptores unitérminos ( $a$ ), bitérminos ( $b$ ), tritérminos ( $c$ ), cuadritérminos, etc. y la aplicación de la siguiente fórmula:

$$t = \frac{a+2b+3c+4d\dots}{a+b+c+d\dots}$$

Las introducciones a los tesauros evaluados no aportaban determinados datos necesarios para poder calcular algunos de los parámetros aquí mencionados, lo cual nos obligó a examinar minuciosamente cada uno de ellos para hacer la estimación de dichas variables. A excepción de la reciprocidad de las relaciones, cuya valoración se hizo mediante muestra aleatoria, los demás datos se consideraron en base a los lenguajes documentales completos para evitar cualquier margen de error a pesar del volumen de términos que los componen. Para hallar las tasas de enriquecimiento y precoordinación, por ejemplo, fue necesario el recuento de todos los descriptores y de todas las relaciones semánticas que comprenden los seis tesauros.

### **5.3. Características de los tesauros analizados**

Pasamos a analizar las particularidades de los tesauros consultados en el trabajo experimental, todos ellos monolingües, tres de los cuales pertenecen al campo de la ciencia y la técnica, y tres a las ciencias sociales y humanidades.

#### **5.3.1. Tesauro ICYT de Biología animal**

- **Composición**

*Contiene una introducción explicativa de las razones que justifican la creación del tesauro, su cobertura temática y la forma en la que se seleccionó y normalizó la terminología. Explica también los símbolos empleados para la representación de las relaciones semánticas entre los términos, así como las formas de presentación, su actualización y cuestiones relacionadas con su tratamiento automático.*

Las formas de presentación de los términos son alfabética, sistemática y permutada.

- **Tamaño**

El tesauro se compone de 3.781 términos, de los cuales 2.782 son descriptores y 999 no descriptores.

- **Relaciones de equivalencia**

Las denominaciones científicas se consideran preferentes, teniendo los términos comunes o vulgares categoría de no descriptores.

La tasa de relaciones de equivalencia es de 0'3, por lo que no alcanza el mínimo estipulado en los criterios de referencia, que señalan como valor más bajo aceptable 0'5, considerando los especialistas más exigentes en este sentido, que el valor deseable debe ser mayor de 1.

- **Relaciones jerárquicas y asociativas**

$$\frac{8.150+971}{2.782} = 0'3$$

- **Reciprocidad de las relaciones semánticas**

En las 27 páginas analizadas, escogidas de forma aleatoria, observamos la existencia de reciprocidad entre todas las relaciones establecidas entre los términos, tanto jerárquicas como asociativas y de equivalencia.

- **Notas aclaratorias, definitorias o de aplicación**

El tesoro tiene un total de 84 notas, lo que supone un 3% del total de descriptores.

- **Niveles jerárquicos**

Se compone de 15 campos temáticos en los que existe un predominio de los niveles jerárquicos 3 y 4, a excepción del campo número 13, destinado a Taxonomía y sistemática, en donde abundan los niveles 7 y 8, llegando a alcanzar en algunos casos hasta 11 niveles de jerarquía descendente.

- **Morfología de las palabras**

Tanto en lo que a forma de las palabras se refiere, como al uso del singular y plural, se atiene a lo establecido en las normas.

- **Aspectos estéticos de composición y tipografía**

Presenta los descriptores en mayúsculas y los no descriptores en minúsculas. De los seis tesauros analizados, éste es el único que representa las relaciones entre términos mediante los símbolos internacionales propuestos por la norma ISO, los demás lo hacen por medio de signos alfabéticos. Ej.:

**Aguila perdicera**

= > HIERAETUS FASCIATUS

**HIERAETUS FASCIATUS**

TM 130401113

= Águila perdicera

< ÁGUILAS

< ACCIPITRIDAE

< FALCONIFORMES

- **Nivel de precoordinación**

$$a = 1.817$$

$$b = 914$$

$$c = 32$$

$$d = 2$$

$$t = \frac{1.817 + 2(914) + 3(32) + 4(2)}{1.817 + 914 + 32 + 2} = 1,36$$

### 5.3.2. Tesauro Electrotécnico

- **Composición**

Se componen de índices alfabético, sistemático y permutado, e incluye un anexo con un diccionario inglés/español.

Es el único lenguaje documental de los analizados que explica en su introducción las posibilidades de compatibilidad con otros tesauros del área o con macrotesauros. Lo hace en un epígrafe en el que señala también la periodicidad aconsejada para su actualización, tanto de la edición experimental como de ediciones posteriores.

- **Tamaño**

Comprende un total 7.764 términos: 6.869 descriptores y 895 no descriptores. El tesauro tiene gran amplitud semántica, incluyendo conceptos muy específicos; ese nivel de concreción nos fue de mucha utilidad al preparar las estrategias de búsqueda, pero

dificultó la localización de información en aquellas ocasiones en las que se necesitaba recurrir a descriptores genéricos, inexistentes en el vocabulario.

- **Relaciones de equivalencia**

La tasa de estas relaciones es de 0'1, cifra muy baja si tenemos en cuenta los valores medios aconsejados.

- **Relaciones jerárquicas y asociativas**

$$\frac{7.796+1.459}{6.869} = 1'35$$

- **Reciprocidad de las relaciones semánticas**

Se procedió al análisis por muestra aleatoria cada once páginas, dada la extensión del tesoro. Se encontraron numerosas irregularidades que mencionamos a continuación. Consideramos que dichas imprecisiones se deben al extenso volumen de términos que recoge y, sobre todo, a que el tesoro constituye una edición provisional, que no ha sido aún revisada.

En el establecimiento de relaciones jerárquicas observamos los siguientes fallos:

El descriptor *anclaje* tiene como término genérico *elementos auxiliares*, que no contiene a su vez *anclaje* como término específico. El mismo caso de falta de reciprocidad se da entre los siguientes descriptores: *andamio* y *material de obra*; *central de gas* y *central térmica de gas*; *conformidad a norma* y *normalización y certificación*;

*convertidor electrónico y dispositivo electrónico*, bajo este descriptor aparece además el término específico *convertidor de fase* en el que no figura como término genérico. En este mismo término aparece la relación con *máquinas*, el cual, realizada la comprobación, se detectó que no lo tiene a él como asociado. Tampoco existe reciprocidad entre *equipos de espuma* y su específico *laza de espuma*, que no existe como descriptor. *Fraguado* tiene como término genérico *material de obra*, pero éste no lo tiene a él como específico. El término genérico *vehículos de tracción eléctricos* incluye como específico *grupo auxiliar*, pero bajo esta entrada no figura ninguna relación de generalidad con aquél; lo mismo ocurre con el descriptor *grupo compresor*. *Guiaondas* cita como genéricos *canal de transmisión* y *circuito de comunicación*, sin embargo, en el primero no está como específico y el segundo no aparece como descriptor. *Normativa de control* remite al término específico *medición*, que no está aceptado como término preferente. *Ruido eléctrico* envía hacia el genérico *transmisión de energía eléctrica*, que no existe como descriptor. *Transporte de energía eléctrica* figura en descendente jerarquía bajo *línea de transporte de energía*, no existiendo reciprocidad en la relación; bajo ese mismo descriptor figuran varios términos específicos que, no lo incluye a él como genérico: *alimentación eléctrica, alta tensión, corriente de línea eléctrica, corriente vagabunda, defecto de línea eléctrica, derivación a tierra, descarga atmosférica, impedancia característica, media tensión, muy alta tensión, poder de cierre en cortocircuito, propagación de la corriente eléctrica, ruido eléctrico, sobre tensión de maniobra, tensión eléctrica, y ultra alta tensión*. *Ventana* tiene como genérico *material de obra*, pero éste no lo tiene a él como específico. *Ventilación doble* remite con un TG a *elementos auxiliares*, pero en esta entrada no figura la existencia de términos específicos; en *ventilación radial* ocurre lo mismo. Por último, *zona neutra* tiene como término genérico *servicio*, pero *servicio* no lo tiene a él como específico.

Se dan algunos casos en los que un descriptor remite a otro más genérico que no aparece en el tesoro como término preferente. Esto ocurre, por ejemplo entre *botón pulsador* y *accionamiento*, entre *magnitudes características* y *tracción eléctrica* y entre *Zimbabwe* y *comunidades españolas*, cuyo error, en este caso resulta evidente.

En cuanto a las relaciones asociativas, se detectaron las siguientes irregularidades:

*Metales* está asociado con *máquinas especiales*, pero éste no está relacionado con *metales*. *Normalización de control* está relacionado con *instrumento-patrón*, relación que no se recoge de forma recíproca, si bien es cierto que la asociación se establece con *normativa de control*, lo cual parece deberse a un error de carácter tipográfico o de cambio de denominación del descriptor a la hora de su inclusión en el tesoro. *Parámetros físico mecánicos* se asocia a *magnitudes de los cables* y a *propiedades de los materiales*, no existiendo reciprocidad entre ellos. Ambos se relacionan, no obstante, con *medida de parámetros físico mecánicos*, lo cual, como en el caso anterior, hace pensar en un posible cambio de denominación del término preferente. *Política de empleo* está relacionado con *seguridad e higiene*, pero en éste no aparece como término relacionado. *Protección auditiva* se asocia a *amortiguador de ruidos* y a *taponos antiruidos*, pero en estas dos entradas no hay reciprocidad con *protección auditiva*. *Sistema eléctrico* está asociado con *rele*, que no aparece en el tesoro; también lo está a los términos *automatización de sistemas* y *calidad de servicio*, pero en estos descriptores no hay relación con *sistema eléctrico*. El término preferente *transporte de energía eléctrica* figura relacionado a *consumo de energía eléctrica*, sin que exista reciprocidad en dicha relación.

- **Notas aclaratorias, definitorias o de aplicación**

El tesoro tiene un total de 256 notas, lo que supone un 3'7% del total de descriptores.

- **Niveles jerárquicos**

Predominan los niveles tres y cuatro, habiendo, en proporción, pocos casos en los que la jerarquía desciende a un quinto nivel.

- **Morfología de las palabras**

- Utiliza sustantivos para representar los descriptores simples y, en el caso de los sintagmáticos, la base es un sustantivo.

- No hemos detectado el uso de la forma invertida de las expresiones.

- Representa los conceptos en singular, excepto cuando el término es más popular o existen diferencias semánticas al usar ambas formas.

- Emplea siglas, que desarrolla mediante notas aclaratorias. Ej.:

**CEDEFOP**

**NA: CENTRO EUROPEO PARA EL DESARROLLO  
DE LA FORMACIÓN PROFESIONAL**

● **Aspectos estéticos de composición y tipografía**

En la presentación alfabética, tanto los descriptores como los no descriptores figuran en mayúsculas y negrita. Al tener el mismo tratamiento tipográfico los términos preferentes y no preferentes, el tesoro no sigue las directrices señaladas al respecto por la norma ISO.

En la parte permutada, consistente en un índice KWIC, se destaca en negrita la palabra que alfabetiza, lo cual supone una ventaja para la visualización de las entradas que lo componen.

Basa el vocabulario en propiedades específicas, procedimientos particulares y materias específicas, razón por la cual hace uso del singular con preferencia al plural, a diferencia de otros tesauros.

● **Nivel de precoordinación**

a= 1.814  
b= 4.314  
c= 954  
d= 61  
e= 6

$$\frac{1.814 + 2(4.314) + 3(954) + 4(61) + 5(6)}{1.814 + 4.314 + 954 + 61 + 6} = 1/9$$

### 5.3.3. Tesoro de Medio ambiente

- **Composición**

Tiene tres formas de presentación: sistemática, alfabética y permutada. Precede al tesoro una introducción explicativa de sus antecedentes, contenido, estructura, criterios de establecimiento de relaciones semánticas, formas de presentación, gestión y utilización.

- **Tamaño**

Se compone de 2.320 descriptores y 608 términos equivalentes.

- **Relaciones de equivalencia**

$$t = \frac{608}{2.320} = 0'26$$

- **Relaciones jerárquicas y asociativas**

La tasa de enriquecimiento de este tesoro es del 2'5%, entrando dentro de los límites aconsejados.

- **Reciprocidad de las relaciones semánticas**

En el análisis que realizamos por muestreo aleatorio no observamos la existencia de irregularidades en lo que a reciprocidad entre las relaciones se refiere, salvo en la

página 67, en donde aparece el término *benzopirenos* en relación con otros descriptores más específicos:

Benzopirenos  
MT 2030 Contaminantes  
BT1 Contaminantes orgánicos  
BT2 Contaminantes químicos

Bajo la entrada *contaminantes químicos* se observa, sin embargo, que no existe ninguna relación recíproca con *contaminantes orgánicos* ni con *benzopirenos*.

● **Notas aclaratorias, definitorias o de aplicación**

Hay un total de 66 notas en el tesoro, lo que supone que tan sólo un 2'8% de los descriptores las incluyen. Se podría justificar un valor tan bajo si se tiene en cuenta que se trata de un tesoro especializado en un área técnica, menos susceptible a la ambigüedad semántica que los destinados a las ciencias sociales. En cualquier caso el índice está muy por debajo del 10% que recomiendan las distintas fuentes.

Las notas responden a la siguiente tipología de uso: definición de conceptos; notas de aplicación, es decir, indicaciones de la forma más adecuada de utilizar un descriptor para indizar o para recuperar información; se emplean, por último, para desarrollar siglas. Ej.:

PLATAFORMA CONTINENTAL  
SN Lecho y subsuelo de las aguas submarinas adyacentes a las costas donde los estados ribereños ejercen derecho de soberanía para la explotación de los recursos.

## EQUIPO DE TRATAMIENTO DE AGUA

SN Se recomienda el uso de descriptores más específicos.

## ONU DI

SN Organización de las Naciones Unidas para el Desarrollo Industrial.

### ● Niveles jerárquicos

La media de niveles oscila entre 5 y 6.

### ● Morfología de las palabras

No hemos detectado la existencia de ningún caso de inversión de términos, respetándose en todas las entradas el orden del lenguaje natural.

Según se indica en la introducción al propio tesoro, se siguen las normas de Unesco en lo que a forma de las palabras se refiere. *Utiliza el singular para representar propiedades específicas, procedimientos particulares, materias específicas, nombres propios, disciplinas, y, en general, conceptos que no pueden ser numerados. Se usa el plural en los descriptores indicativos de las propiedades genéricas, los materiales genéricos, los objetos, y, en general, conceptos que pueden ser numerados*<sup>11</sup>. Se respeta igualmente la forma sustantiva de las palabras. Ej.:

---

<sup>11</sup>UNESCO; ISO. *Principes directeurs pour l'établissement et le développement de thesaurus monolingues*. Paris: UNESCO; Genève: ISO, 1981. (Norma ISO 2788-1981).

factor climático	invernaderos
ingeniería	montes
productividad	productos agrícolas

El tratamiento de las siglas es diverso, del análisis del índice alfabético del tesoro deducimos su empleo de la siguiente forma: a) se desarrolla en nota su significado; b) se toman las siglas de la institución como término descriptor y se considera la forma desarrollada como no descriptor. Por último, se trata como término preferente la forma desarrollada, en cuyo caso las siglas pasarán a la categoría de no descriptores, con el consiguiente reenvío hacia la forma aceptada. Ej.:

UNICEF  
SN Fondo de las Naciones Unidas para la Infancia.

OMS  
UF Organización Mundial de la Salud

EURATOM  
USE Comunidad europea de la energía atómica

- **Aspectos estéticos de composición y tipografía**

Utiliza signos alfabéticos normalizados, en inglés, para indicar el tipo de relación establecida entre los términos.

Cada una de las listas que componen el tesoro está impresa en papel de diferente color, lo cual es una ayuda para localizar la información. No existe, sin embargo, diferenciación tipográfica entre descriptores y no descriptores, ambos se expresan en minúscula y en negrita.

● **Nivel de precoordinación**

$$a = 870$$

$$b = 1.400$$

$$c = 94$$

$$d = 9$$

$$t = \frac{870 + 2(1.400) + 3(94) + 4(9)}{870 + 1.400 + 94 + 9} = 1,68$$

**5.3.4. Tesauro "Mujer"**

● **Composición**

Lo componen cuatro partes: a) un índice alfabético que se limita a listar descriptores y no descriptores sin indicación detallada de las relaciones semánticas entre los términos; b) un esquema jerárquico que comprende todos los campos del tesauro; c) un índice sistemático que incluye atípicamente no descriptores y notas aclaratorias y d) un índice permutado Kwic. En contra de lo que aconsejan las normas, sitúa las relaciones de equivalencia de cada descriptor después de las jerárquicas y asociativas, por ejemplo:

**AUDIOVISUALES**

TE VIDEO

FOTOGRAFÍA

TR CINE

UP MEDIO AUDIOVISUAL

- **Introducción**

Aunque no especifica periodicidad, prevé actualizaciones periódicas que incluyan posibles candidatos a descriptor.

- **Tamaño**

El tesoro está compuesto por 1.118 términos, de los cuales 714 son descriptores y 404 son términos equivalentes.

- **Relaciones de equivalencia**

$$t = \frac{404}{714} = 0'56$$

- **Relaciones jerárquicas y asociativas**

$$t = \frac{282 + 664}{714} = 1'32$$

Dado que la tasa aconsejada es de 2 a 5, el índice de relaciones en este tesoro resulta muy baja.

● **Reciprocidad de las relaciones semánticas**

En el análisis de la muestra preparada para comprobar la reciprocidad, se observó que en las relaciones de equivalencia existe una falta de correspondencia en los siguientes casos:

Los términos no descriptores *Prestaciones por incapacidad laboral transitoria, distribución de responsabilidades familiares, pobres, conflictos armados, programas de salud y trabajadoras madres de familia*, que figuran precedidos del signo UP en el índice sistemático, han sido omitidos en los índices permutado y alfabético.

*Directoras cinematográficas e informáticas* no figuran ordenados alfabéticamente.

Al realizar la comparación entre los términos específicos (TE) y los genéricos (TG), observamos la existencia de algunas incorrecciones al desarrollar los específicos, así como ciertos fallos en la organización alfabética de los mismos:

El TE *exiliadas*, dependiente de *migración*, no figura en orden coherente. Los términos específicos que derivan del genérico *protección social* no siguen un orden alfabético, estando además sin establecer las relaciones del descriptor *servicios sociales*. En el TG *iniciativas locales de empleo* falta por desarrollar el TE *autoempleo*. En el TG *cargos públicos* existe el TE *asociacionismo*, que tampoco está desarrollado; por último, el descriptor genérico *voto femenino*, perteneciente al campo temático anterior incluye *electorado femenino*, pero omite la explicación de sus relaciones.

En la página 98 se repite el término *Enfermedades mentales*.

Bajo el descriptor genérico *Profesionales de las artes* no se incluye el descriptor específico *Trabajadoras de artes gráficas*, que se considera parte de este campo jerárquico, y sin embargo se desarrolla el término *Músicas*, que no se incluía en dicha cadena jerárquica.

Por lo que a relaciones asociativas se refiere, no se han detectado irregularidades, salvo alguna de carácter tipográfico.

En el TG *Bodas* aparece el término relacionado *Nupcialidad*, pero en el término genérico *Nupcialidad* no existe la relación recíproca. No hay reciprocidad tampoco entre los descriptores *Centros de enseñanza y Educación*.

- **Notas aclaratorias, definatorias o de aplicación**

El tesoro tiene un total de 123 notas, lo que supone un 17% del total de descriptores.

- **Niveles jerárquicos**

El máximo de niveles jerárquicos es de seis, siendo la media de tres, la más utilizada en el conjunto del tesoro. Por ejemplo, en el descriptor *trabajadoras*, que aparece en las páginas 61 y 62.

- **Morfología de las palabras**

Se utiliza el género femenino plural y no el masculino, de acuerdo con las características del Centro de documentación y el tipo de documentación que recoge, centrada en la problemática de la mujer.

- **Aspectos estéticos de composición y tipografía**

Descriptores y no descriptores se escriben en mayúsculas, destacándose los términos aceptados en negrita. En el índice permutado -Kwic-, se resaltan en negrita las entradas que alfabetizan.

- **Nivel de precoordinación**

$$a = 362$$

$$b = 324$$

$$c = 32$$

$$t = \frac{361 + 2(321) + 3(32)}{361 + 321 + 32} = 1'5$$

### **5.3.5. Tesauro ISOC de Psicología**

El tesauro, editado en 1995, constituye la versión actualizada del publicado en 1992 por iniciativa del ISOC, con la intención de servir de instrumento de indización a los Centros de Documentación y Bibliotecas especializadas en Psicología y Ciencias afines.

## ● Composición

Se compone de cuatro partes: un índice alfabético con todas las relaciones de equivalencia, jerárquicas y asociativas de cada descriptor; un índice sistemático que agrupa los descriptores en campos temáticos; un índice permutado y un anexo. Incluye además un apéndice que comprende la clasificación para el área de psicología en la base de datos PSEDISOC, que está basada en la utilizada por el Psychological Abstracts y PsycINFO (American Psychological Association) en su versión de 1988.

El índice permutado comprende no descriptores, pero no remiten al descriptor aceptado, por lo que el usuario no puede conocer el campo al que pertenece, a menos que consulte, además, la parte alfabética del tesoro.

Contiene una introducción en la que se informa al usuario de su ámbito de aplicación y de sus características.

La codificación es numérica. El código se sitúa detrás del descriptor en los índices alfabético estructurado y permutado, y va precediéndolo en el índice sistemático.

El tesoro es de carácter monodisciplinar, incluyendo vocabulario de campos auxiliares para cubrir aspectos económicos, jurídicos, culturales, artísticos, psicológicos y sociales.

Para su elaboración se combinaron los métodos inductivo y deductivo, utilizando como base terminológica las palabras clave que sirvieron para la indización de los artículos de revistas que venía realizando el ISOC desde el año 1975 en que se creó la Base de Datos de Psicología, denominada en la actualidad PSEDISOC. A estas palabras

clave se sumaron los términos utilizados por los usuarios para solicitar documentación del Centro, respondiendo de este modo el vocabulario a necesidades de información reales. La terminología se completó con el vaciado de diccionarios, léxicos, planes de estudio y tesauros especializados o afines a la materia, existentes en España.

- **Tamaño**

El tamaño del tesoro es de 4.400 términos de los cuales 2.788 son descriptores y 1.612 no descriptores. La parte del anexo contiene 690 términos, de los cuales 423 son no descriptores y 359 descriptores, que se agrupan en 17 campos temáticos.

- **Relaciones de equivalencia**

$$t = \frac{1.612}{2.788} = 0'57$$

- **Relaciones jerárquicas y asociativas**

No incluye relaciones polijerárquicas, siendo la tasa de enriquecimiento del 2'9%:

$$t = \frac{5.249 + 2.868}{2.788} = 2'9$$

- **Reciprocidad de las relaciones semánticas**

Todas las relaciones que se establecen son recíprocas.

- **Notas aclaratorias, definitorias o de aplicación**

Del total de 2.788 descriptores, 732 tienen notas aclaratorias, siendo el porcentaje del 26%.

Las notas se utilizan básicamente para definir conceptos especializados. Ej.:

### MEDITACIÓN

NA Reflexión sostenida en estado de relajación, en el cual una persona intenta lograr la integración de sus sentimientos, emociones, actitudes e ideas.

- **Niveles jerárquicos**

El máximo de niveles jerárquicos del tesoro es de ocho.

- **Morfología de las palabras**

Su uso se adecúa a lo que dictan las normas. La forma plural se destina a temas concretos: *psicoterapeutas*, y el singular a temas abstractos: *valoración*.

La base del descriptor es sustantiva. Utiliza el sustantivo o formas sustantivadas, sustantivos más adjetivos y sustantivos unidos por preposición.

En ningún caso hemos observado el empleo de inversión de términos, se respeta el orden natural de las expresiones.

● **Aspectos estéticos de composición y tipografía**

En la parte alfabética, tanto descriptores como no descriptores se escriben en mayúsculas, pero éstos últimos están destacados tipográficamente en negrita. En el índice permutado, sin embargo, no existe diferenciación tipográfica entre términos preferentes y no preferentes, lo que, unido a la ausencia de reenvío de términos no aceptados a los que sí lo están, resta utilidad a dicho índice. En la presentación sistemática los distintos niveles jerárquicos se representan mediante sangrado a la derecha, haciendo preceder, además, a cada descriptor de puntos suspensivos que facilitan la visualización de dichos niveles.

● **Nivel de precoordinación**

$$a = 1.231$$

$$b = 1.358$$

$$c = 186$$

$$d = 12$$

$$e = 1$$

$$t = \frac{1.231 + 2(1.358) + 3(186) + 4(12) + 5(1)}{1.231 + 1.358 + 186 + 12 + 1} = 1,63$$

Nota.- Para hacer las valoraciones cuantitativas de este tesoro no se han contabilizado los términos componentes del anexo, porque consideramos que en su composición se ha seguido el mismo criterio que en el resto del tesoro, y por la amplitud que ya de por sí tiene la parte nuclear del mismo, cuyas entradas por sí solas permiten realizar una evaluación exhaustiva.

### 5.3.6. Tesoro de Asuntos sociales

- **Composición**

Se compone de listas alfabética, sistemática y permutada, y se completa con una introducción explicativa de la intencionalidad del tesoro y de las razones que motivaron su creación. Se informa en ella al usuario sobre las siguientes cuestiones: ámbito de aplicación, metodología, estructura y presentación.

De la necesidad de actualización dan fe algunas entradas del tesoro tales como: *Alemania R.D.*, *Alemania R.F.* La razón por la que las revisiones y puestas al día de éste y otros tesoros no se hacen con la periodicidad deseable es que no todas las instituciones cuentan con recursos materiales y humanos para dicha actualización, debido, en buena parte, a la escasez de medios con que se dota, en ocasiones, al quehacer bibliotecario.

- **Tamaño**

El tesoro se compone de 1.710 descriptores y 429 términos no descriptores.

- **Relaciones de equivalencia**

$$t = \frac{429}{1.710} = 0,25$$

● **Relaciones jerárquicas y asociativas**

$$t = \frac{1860 + 511}{1.710} = 1'38$$

● **Reciprocidad de las relaciones semánticas**

Aunque en el tesoro no existen muchas irregularidades en este sentido, detectamos la existencia de falta de correspondencia entre algunos términos en lo que a relaciones asociativas y de equivalencia se refiere. Por ejemplo: el término genérico *centros de documentación* tiene relaciones asociadas con *difusión de la información* y con *documentación*, sin embargo, en la entrada correspondiente a *difusión de la información* no aparece la relación recíproca con *centros de documentación*. Se da el mismo caso entre *centros de rehabilitación*, conectado a *minusválidos*, no existiendo reciprocidad a la inversa. Otros descriptores como *centros especiales de empleo* o *centros ocupacionales* tienen asociado el descriptor *minusválidos*, pero éste último no cuenta entre sus relaciones con estos dos descriptores que deberían estar asociados a él. El término preferente *información* está relacionado con *ciencias de la información*, en cuya entrada no figura el término relacionado *información*.

En cuanto a las relaciones de equivalencia se dan los siguientes casos:

*Deficiencia desfiguradora* aparece como equivalente de *deformidad*, pero no existe ninguna entrada en la que figure dicho término como no descriptor. En la entrada genérica *deficiencia mental grave* se incluye la relación equivalente con *retraso mental grave*, sin que exista entrada recíproca en el tesoro. Por último, las siglas *OCDE* se aceptan como descriptor, relacionando con un UP el término no aceptado *Organización*

*de Cooperación y Desarrollo Económicos*, sin embargo este último término no preferente aparece como reenvío.

- **Notas aclaratorias, definatorias o de aplicación**

Se limitan en este tesoro a definir el ámbito de aplicación del vocabulario. Existen un total de 24, es decir, el 1'4%, el porcentaje más bajo de los tesauros analizados.

- **Niveles jerárquicos**

La tasa máxima es de cinco niveles jerárquicos, aunque predominan los campos con tres y cuatro niveles.

- **Morfología de las palabras**

Como se señala en la propia introducción, el tesoro se ciñe a las reglas establecidas internacionalmente, en lo que a morfología del descriptor se refiere.

En cuanto a las siglas, se utiliza en unos casos la forma desarrollada y en otros la abreviada. Cuando se adopta ésta última como descriptor, se hace un reenvío hacia la forma completa, que considera como término no descriptor.

- **Aspectos estéticos de composición y tipografía**

Los descriptores están escritos en negrita, lo cual facilita su distinción de los no descriptores, que no aparecen destacados, sin embargo, tanto los términos preferentes

como los no preferentes figuran en minúscula, lo cual no beneficia la localización de los descriptores.

Los términos no aceptados van acompañados de las siglas ND (No Descriptor), innecesaria a nuestro juicio, puesto que, como en todos los tesauros, dichos términos llevan la indicación USE, que remite al término aceptado. La inclusión de la sigla mencionada supone además un código añadido que ha de interpretar el usuario.

● **Nivel de precoordinación**

a= 857  
b= 780  
c= 67  
d= 6

$$t = \frac{857 + 2(780) + 3(67) + 4(6)}{857 + 780 + 67 + 6} = 1'54$$

**5.4. Características de las bases de datos consultadas**

**5.4.1. Base de datos ICYT de Biología animal**

La información relativa a Biología animal se encuentra recogida en la base de datos de Ciencia y Tecnología del ICYT, que distribuye el CINDOC. Cubre un amplio campo de materias y tiene además conexión con otras ramas de la ciencia tales como veterinaria y producción animal.

La base se nutre básicamente de publicaciones seriadas de carácter científico no de divulgación, y, en menor proporción, de monografías, boletines e informes. Esta información cubre todo lo publicado en España en el período de 1979 a 1997.

Se puede consultar en línea o en CD-Rom. Los discos se actualizan cada dos años, mientras que la puesta al día en línea es diaria. El CD-Rom lo comercializa Micronet, la documentación en línea, como decíamos anteriormente la distribuye el CINDOC.

Es posible recuperar por todos los campos, tanto en línea como en CD-Rom. Dichos campos son: número de registro, autor, título, título original, lugar de trabajo, nombre de la revista, publicación, ISSN, tipo de documento, idioma, clasificación, descriptores, descriptores secundarios, topónimos, identificadores y período histórico.

Al tratarse de un área del conocimiento que abarca varias disciplinas, el nivel de indización no puede ser muy especializado, no obstante, se tiende a profundizar lo más posible en el momento de caracterizar los documentos. El lenguaje controlado empleado para este proceso es el *Tesaurus ICYT de Biología animal*, que analizamos en este mismo capítulo.

El horario de acceso a la base es de 8 de la mañana a 8 de la tarde. Se puede obtener ayuda para su consulta, aunque tenemos que decir a este respecto que este servicio, al menos en lo que respecta a nuestra experiencia, es muy precario debido a la escasez de terminales disponibles para el usuario, así como las dificultades técnicas encontradas en su utilización. Para realizar las consultas que desarrollamos en el capítulo sexto de este trabajo, no tuvimos otra alternativa que usar el CD-Rom, porque el acceso

en línea no nos era posible. El interfaz es en esta modalidad sencillo de utilizar, y ofrece la posibilidad de utilizar recursos como el álgebra de Boole y otro tipo de enlaces sintácticos para conformar las estrategias. Las primeras pantallas que encontramos al realizar la conexión se representan en las figuras 39, 40 y 41.

#### **5.4.2. Base de datos PIE (Programa de Investigación y Desarrollo Electrotécnico).**

La base depende de la Asociación de Investigación Industrial Eléctrica (Asinel) y BASIS es el distribuidor que la soporta. Se puede consultar en línea durante las veinticuatro horas del día, vía modem.

Se compone de 1.360 documentos, todos ellos contienen proyectos de investigación eléctrica. La base se creó en 1988, no existiendo antecedentes de una similar hasta esa fecha.

La actualización de los datos se realiza mensualmente. Dichos datos se distribuyen en soporte magnético (pack) dos veces al año a las empresas eléctricas, por lo que las consultas directas a la base de Asinel son en muchos casos innecesarias y, en cualquier caso, difícilmente controlables.

El nivel de indización de los documentos es muy específico, de acuerdo con la naturaleza de la materia que tratan. Se utiliza para la normalización el *Tesoro Electrotécnico*, cuyas características se analizan en este mismo capítulo.

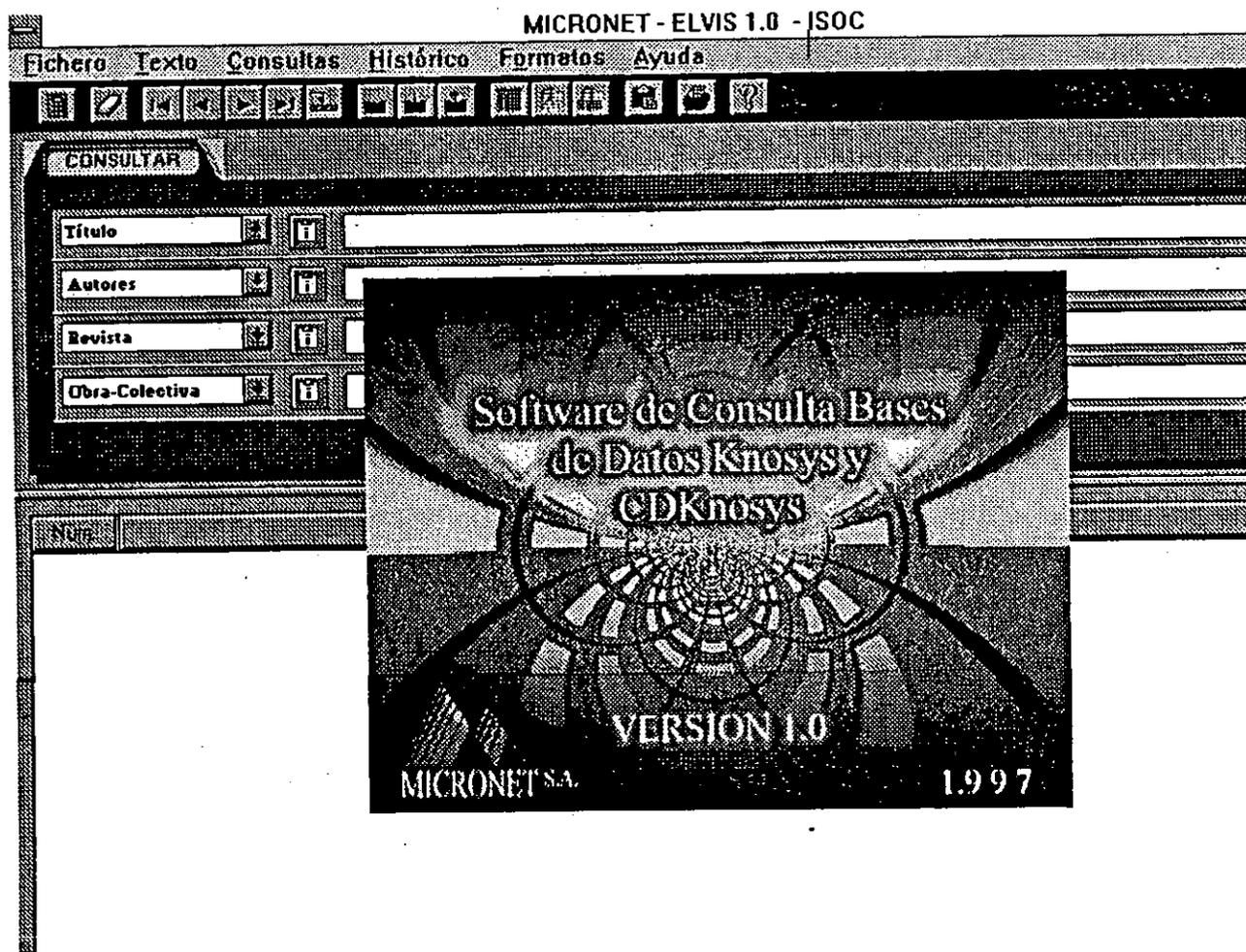


Figura 39. Primeras pantallas del menú de las bases de datos del CSIC

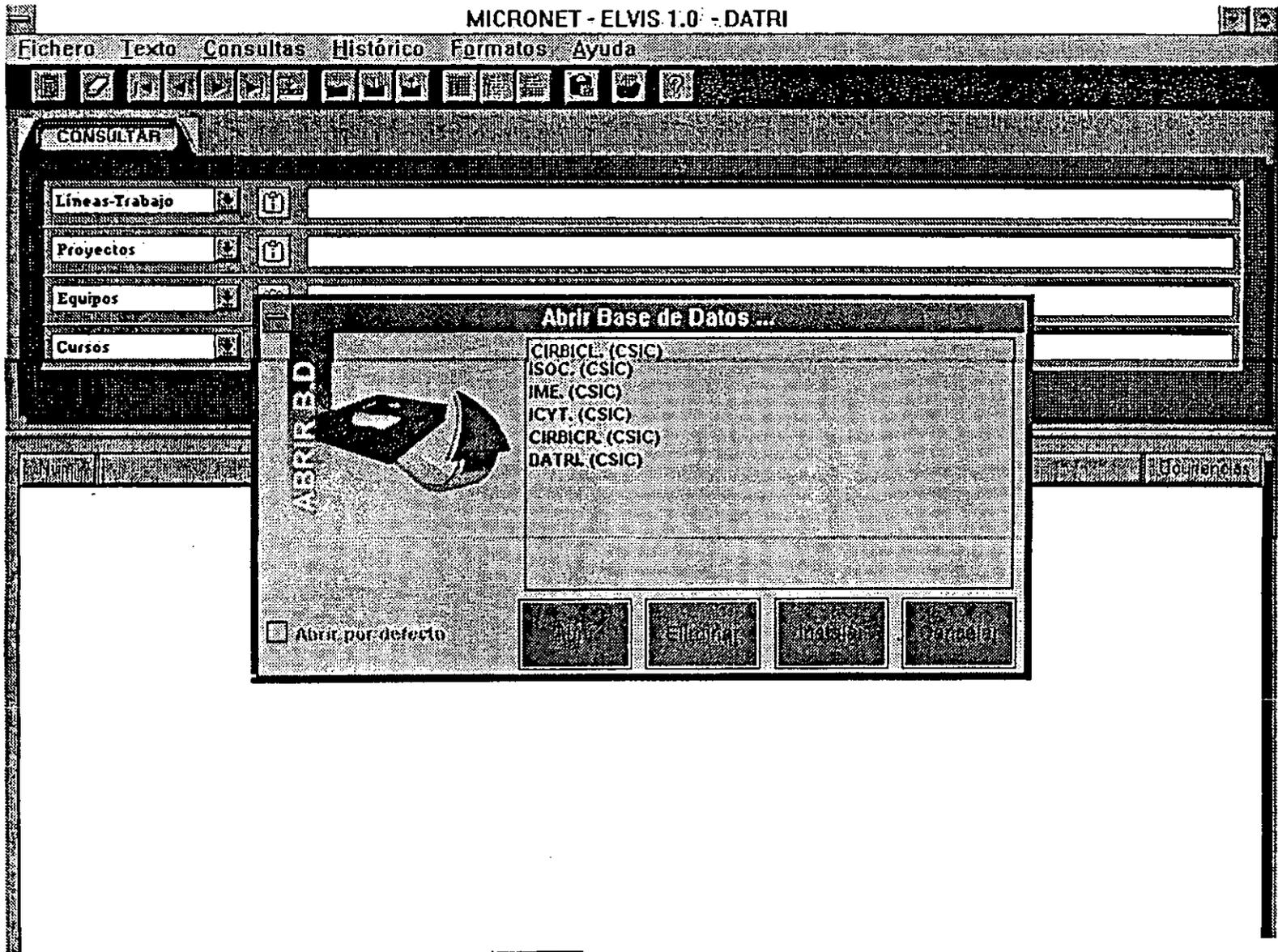


Figura 40

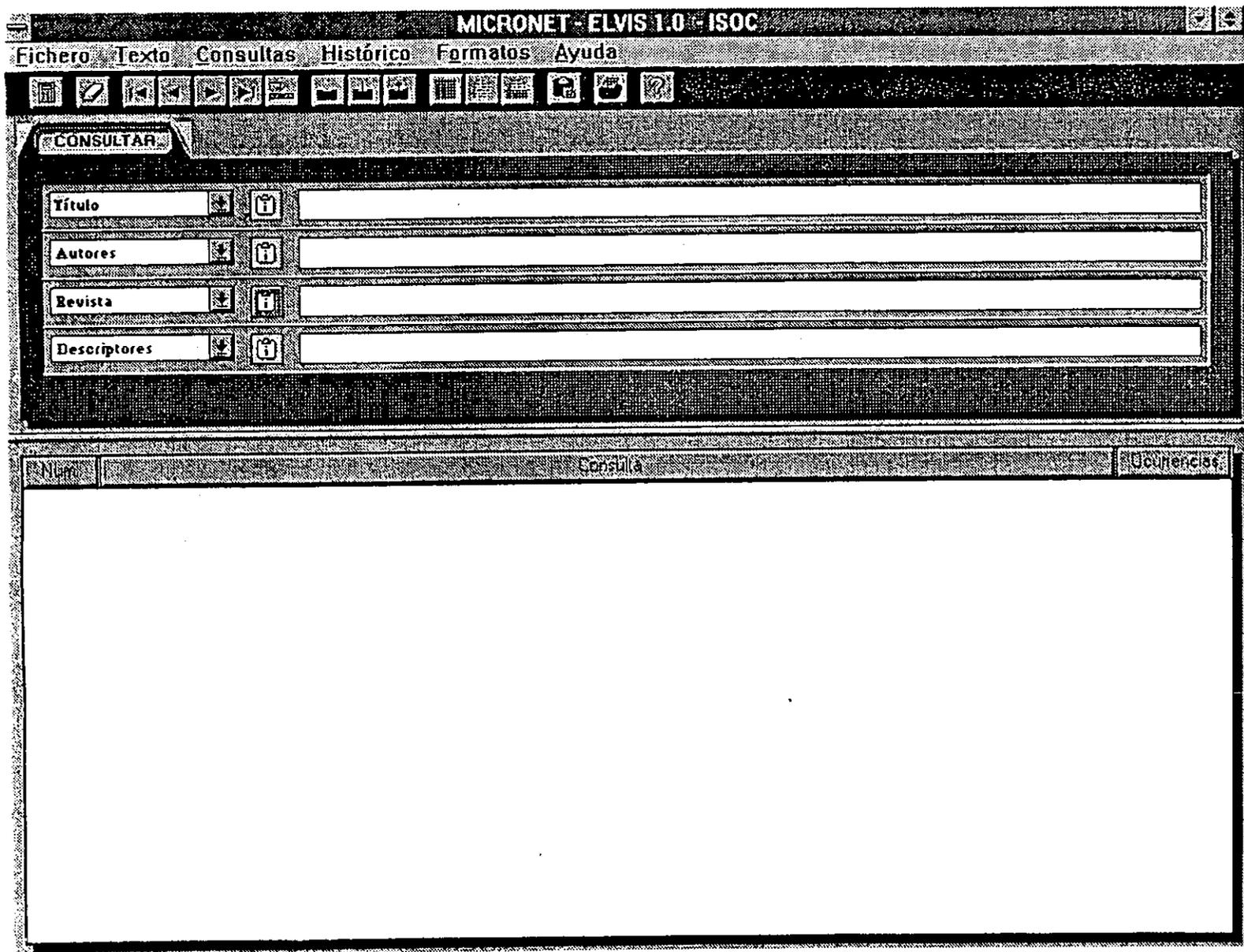


Figura 41

Al establecer la conexión, aparecen en pantalla los siguientes datos:

Primera pantalla:

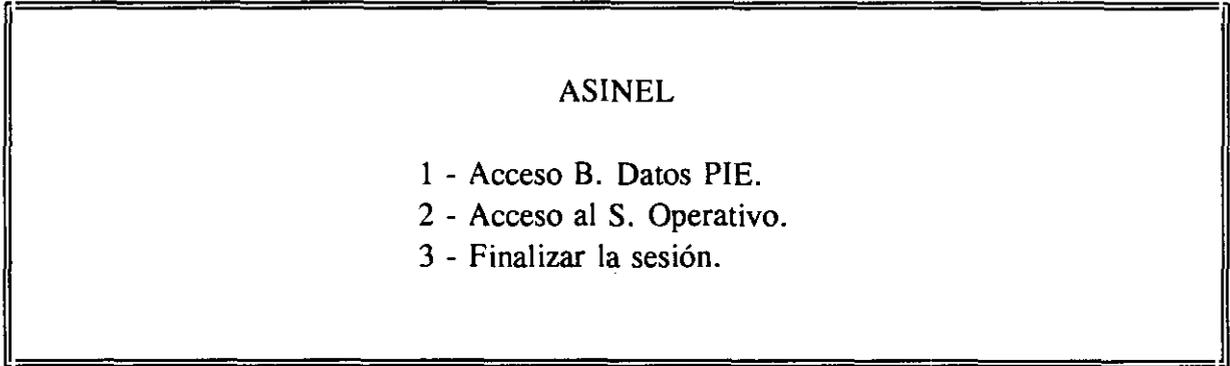


Figura 42

Segunda pantalla:

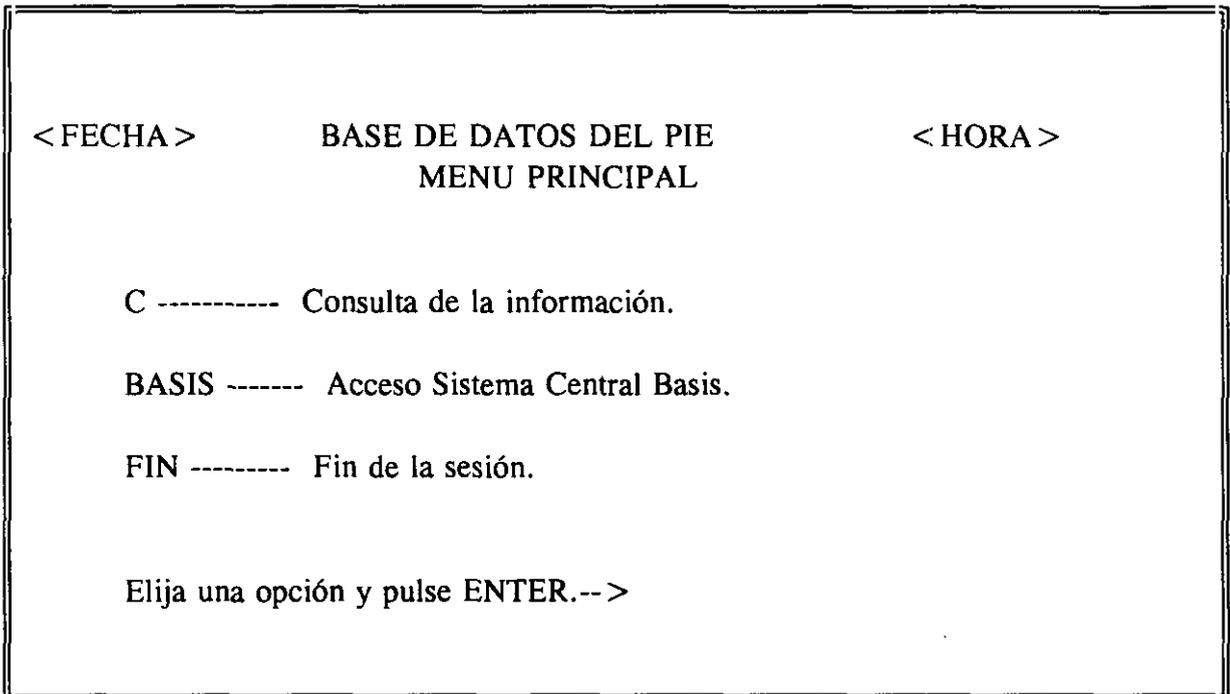


Figura 43

Tercera pantalla:

<FECHA> BASE DE DATOS DEL PIE <HORA>  
N. DOCUMENTOS SELECCIONADOS.--  
CONSULTA B. DATOS

NP - Num. PIE	SI - Signatura.
NR - Num. Registro	EP - Estado del proyecto
TI - Título	EV - Evaluación
EM - Empresas	PR - Presupuesto
EC - Entidades Colab.	DG - Descriptores Geog.
FI - Fecha de Inicio.	DT - Descriptores Temát.
FF - Fecha Final.	RO - Resumen de objetivos.
AU - Autores.	RE - Resumen.

LIST - Subconjunto búsqueda	NEW - Inicializat Conj.
MENU - Acceso a Menú Principal	FIN - Fin sesión
VIS - Visualización Document.	OUTP - Impresión Subc.
REDU - Visualización Reducida	BASIS - Acceso a Basis.

Elija opción y pulse INTRO

Figura 44

Al consultar se deben escribir en la pantalla los términos de búsqueda puesto que, por el momento, el tesoro no está automatizado. Es posible utilizar los operadores booleanos y la búsqueda por truncado, siendo el tiempo de respuesta inmediato. El sistema detecta los errores sintácticos cometidos en la formulación de las estrategias.

No hay limitación de sentencias a unir, ni tampoco a exportar, lo cual constituye una gran ventaja con respecto a otros sistemas. Esta base ofrece la posibilidad de visualizar todos los campos o sólo algunos de ellos.

Cuando la exportación de los registros se hace en papel, los descriptores, o las palabras clave situadas en los campos recuperados aparecen destacados en cursiva, lo cual supone una gran ayuda para el análisis de los resultados.

En el capítulo cuarto de esta tesis incluimos un modelo de registro de PIE en el que pueden observarse los puntos de acceso al documento, siendo todos ellos recuperables.

Es destacable la facilidad de acceso a esta base de datos, así como el apoyo recibido para su consulta.

#### **5.4.3. Base de datos Biblioma, del Centro de Documentación del Ministerio de Medio Ambiente**

El centro de documentación de Medio Ambiente consta de seis bases de datos: *Biblioma*, compuesta fundamentalmente por monografías; *Lema* que abarca todo lo relativo a legislación en materia de medio ambiente; *Audima* que recoge información audiovisual; *Pima* dedicada a proyectos de investigación; *Rema*, que se nutre de revistas sobre la materia del centro; y *Ema*, que recopila estudios y literatura gris en general, consultable en sala o en microficha.

De todas estas bases tomamos para nuestra experimentación *Biblioma*, que no pudimos consultar en línea por encontrarse el centro de documentación en período de traslado, por lo que decidimos trabajar con el CD-Rom que distribuye el Centro de publicaciones del Ministerio de Medio ambiente. El disco contiene un total de 26.362 registros que cubren el período de 1976 a 1994. Recoge las siguientes materias: Medio

natural, asentamientos, impacto ambiental, política ambiental, derecho y legislación, gestión ambiental, medio social, actividad económica y Ciencia y tecnología.

La actualización de los datos se realiza habitualmente a diario, y el CD-Rom se edita anualmente.

En el CD-Rom todos los campos son recuperables y, por tanto, exportables a papel o a disquete. Existe, no obstante, limitación de sentencias a unir o exportar, siendo el máximo de 9.

Para la normalización de los documentos se utiliza el *Tesoro de Medio ambiente*, cuyas características han sido descritas en este capítulo. Se puede acceder al tesoro de forma automática, lo cual resulta beneficioso para la consulta. Detectamos la falta de correspondencia entre algunos descriptores del tesoro manual y el introducido en la base, viéndonos obligados, en algunas consultas, a modificar las estrategias preparadas originalmente.

No podemos representar aquí el menú de apertura de la base, ya que, desde el CD-Rom, el software de la base -Micronet- no permite su impresión en papel ni su exportación a disquete.

#### **5.4.4. Base de datos del Centro de Documentación de la Mujer**

La base depende del Instituto de la Mujer, en conexión con el servidor de la red del mencionado Instituto.

Contiene 12.700 documentos, todos ellos referidos a la mujer, desde el punto de vista sociológico. El centro recoge libros, informes, actas, literatura gris, vaciado de artículos de revistas, analíticas de capítulos de libros, estadísticas, investigaciones, etc.

La actualización de los datos se realiza cada dos o tres días, volcándose los datos del fichero temporal de carga a la central.

La base cubre el período de 1980 a la actualidad, si bien se realizó un expurgo de información relacionada con cocina, folklore, etc., procedente de los fondos de la Sección Femenina que se habían incorporado a dicho Centro.

El nivel de indización de los documentos es medio, dado el carácter interdisciplinar de las materias relacionadas con el tema central de la base. Para el proceso de normalización se utiliza el *Tesauro "Mujer"*.

Al consultar se deben escribir en la pantalla los términos de búsqueda puesto que, por el momento, el tesauro no está incluido en el programa.

Se pueden unir y exportar sentencias sin limitación, lo cual supone una gran ventaja para el usuario. Esta base ofrece, además, la posibilidad de recuperar por todos los campos, unos tienen ficheros inversos, no secuenciales, por lo que permiten una ágil

recuperación. Incluso el campo resumen, aunque no tiene asociado índice de interrogación, tarda menos de un minuto en encontrar una cadena de caracteres solicitada.

El horario de acceso a la base es de 9 a 2 horas, por las mañanas y de 4 a 6 horas de la tarde. Al establecer la conexión, aparecen en pantalla los siguientes datos:

Primera pantalla:

APLICACIÓN DOCUMENTAL	CHEMDATA - INFOTEX
-----	
CONSULTA - MENU PRINCIPAL	
-----	
<u>Bases de datos disponibles:</u>	
1 - ARTÍCULOS DE REVISTA	
2 - MONOGRAFÍAS	
3 - ARCHIVO CENTRAL	
.	
.	
.	
Elija la opción deseada:	
-----	
FIN = Salir de CONSULTA	

Figura 45

Segunda pantalla:

<b>APLICACIÓN DOCUMENTAL * CHEMDATA - INFOTEX</b>	
<b>BASE DE DATOS: ARTÍCULOS DE REVISTA - CONSULTA SENCILLA</b>	
<b>Campos de consulta:</b>	
<b>1 - Autor Personal</b>	<b>7 - Descriptores</b>
<b>2 - Autor Secundario</b>	<b>8 - Lengua</b>
<b>3 - Autor Corporativo</b>	<b>9 - País</b>
<b>4 - Congreso</b>	
<b>5 - Título</b>	
<b>6 - Tipo de Documento</b>	
<b>CAMPO:</b>	<b>\$0</b>
<b>CONSULTA:</b>	
<b>CM= Cons. Múltiple LA= Listar F. Aut. T= Listar Tesouro</b>	
<b>V= Volver M.</b>	

Figura 46

Tercera pantalla:

APLICACIÓN DOCUMENTAL * CHEMDATA - INFOTEX	
-----	
BASE DE DATOS: ARTÍCULOS DE REVISTA - CONSULTA COMPUESTA	
-----	
\$1	9 respuesta(s) para: AUTORPER=RUÍZ*
\$2	50 respuesta(s) para: DES=ENSEÑANZA PRIMARIA Y EDUCACIÓN ESPECIAL
\$3	300 respuesta(s) para: LENGUA=SPA
Elija la opción deseada: _	
-----	
CS= Cons. Simple CM= Consulta múltiple CH= Consulta Histórico H=Histórico IH=Inic. Histórico L= Listar LC= Listar Campos Sec.	

Figura 47

Se pueden emplear los siguientes recursos: truncamientos a izquierda y derecha, operadores booleanos, y lo que denominan en el propio centro máscaras, que consisten en un punto o más que se pueden escribir en sustitución de un carácter.

Cuando se realiza la exportación de los registros, tanto en papel como en disquete, los descriptores, o las palabras clave situadas en los campos recuperados no aparecen destacados tipográficamente, lo cual dificulta el análisis de los resultados.

La consulta en esta base es necesariamente en línea, siendo destacable la facilidad de acceso a la misma, así como el apoyo recibido durante el desarrollo de las búsquedas.

#### **5.4.5. Base de datos PSEDISOC del CSIC**

La denominación propiamente dicha de la base es ISOC, fichero PSEDISOC. *Depende, como se deduce de su nombre, del ISOC. La base forma una unidad con todas las demás que componen el CSIC, de tal manera que para consultar exclusivamente los documentos relacionados con psicología, nos ceñimos a realizar las búsqueda en la clasificación número 65, que realizaron en la mencionada base de datos adoptando la utilizada por el Psychological Abstracts en versión de 1988.*

El distribuidor de la base es, como es natural, el CSIC. Se puede consultar en línea de 9 de la mañana a 8 de la tarde.

Utiliza el lenguaje de interrogación BASIS. Las primeras pantallas que aparecen al realizar las consultas son las mismas que las relativas a la base de datos de Biología animal comprendida, al igual que ésta, entre las del CSIC (figuras 39 40 y 41).

Contiene 18.089 registros sobre psicología y ciencias afines; la tipología documental se ciñe a artículos de revistas especializadas editadas en España, y en algunos casos, a actas de congresos.

La base cubre el período de 1975 a la actualidad, con una laguna que se sitúa entre los años 1979 y 1980. La información generada a lo largo de esos dos años no se

incluyó porque la falta de tiempo impedía tener actualizada la base con documentos fechados con posterioridad que iban entrando en la base, tomándose la decisión de introducir los nuevos e ir incorporando los de dichos años poco a poco. La actualización se realiza diariamente.

Los campos recuperables son: número de registro, autor, título, título original, lugar de trabajo, nombre de la revista, publicación, congreso (nombre del), ISSN, tipo de documento, idioma, clasificación, descriptores, descriptores secundarios, topónimos, identificadores, período histórico y palabras del resumen<sup>12</sup>.

Se utiliza el *Tesaurus ISOC de Psicología* para la indización de documentos, que se realiza a un nivel de especificidad alto. Aunque el límite preestablecido es de 8 a 10 descriptores por documento, en ocasiones se trabaja sin limitación en este sentido.

El sistema ofrece la posibilidad de truncar por prefijo y sufijo, el uso de adyacencias, de operadores booleanos, y la utilización de enlaces de proximidad, que los propios especialistas consideran de complicado empleo. No hay limitación de sentencias a unir, ni tampoco a exportar y el tiempo de respuesta en teoría es inmediato, aunque depende de la complejidad de la pregunta.

Cuando se exportan los registros, ni los descriptores, ni las palabras clave situadas en los campos recuperados aparecen destacados, lo cual hace más difícil su localización en las referencias.

---

<sup>12</sup>La base tiene campo de resumen, en el que se incluyen los del propio autor del documento.

No es posible el acceso en línea a la base sin ayuda de personal especializado, porque requiere el aprendizaje previo del lenguaje de recuperación, pero existe la posibilidad de realizar la consulta en CD-Rom.

#### **5.4.6. Base de datos de la Biblioteca del Instituto de Migraciones y Servicios sociales.**

La base depende del Instituto de Migraciones y Servicios Sociales, del Ministerio de Asuntos sociales. La Biblioteca cubre un período de 25 años, comenzó a funcionar como *Biblioteca de Servicios sociales* en 1972, y el tratamiento informático se comenzó a realizar en 1991.

El proveedor de la base de datos es la empresa Chem-data. Se puede consultar en línea durante las veinticuatro horas del día, vía modem.

Contiene 21.582 registros, procedentes de todo tipo de documentos: revistas, monografías, informes técnicos, actas de congresos y conferencias, estadísticas, proyectos de investigación, etc. Las materias de este fondo documental se agrupan en 25 campos temáticos.

El centro atiende un promedio de 9.000 consultas por año, siendo el horario de acceso al público de 9:00 a 14:00 horas.

El nivel de indización es medio, de acuerdo con las características de heterogeneidad y, en muchos casos, de generalidad de la información que se recibe en la biblioteca.

El único útil que se emplea para la normalización del tratamiento documental es el *Tesouro de Asuntos sociales*.

La actualización de los datos se realiza dos o tres veces al mes.

Al consultar se pueden escribir en la pantalla los términos de los que se compone la estrategia, o bien, se toman del campo de descriptores, dado que el tesouro empleado para el tratamiento documental está introducido en el programa automático, por lo que aparece en el menú de consulta. Permite utilizar los operadores booleanos y la búsqueda por truncamiento, siendo el tiempo de respuesta inmediato. El sistema detecta, además, los errores sintácticos cometidos en la formulación de las estrategias, e incluye otro recurso que permite arrastrar todos los términos específicos dependientes jerárquicamente de cualquier término genérico que figure en el tesouro; el símbolo utilizado para este procedimiento es [#], que se sitúa precediendo al término genérico en el momento de realizar la estrategia de búsqueda. Ej.:

# JÓVENES, abarca a JÓVENES INADAPTADOS y  
JÓVENES MARGINADOS

Los campos recuperables de esta base de datos son: número de referencia, encabezamiento principal, encabezamiento secundario, título, autor, descriptores, identificadores, descriptores geográficos, lengua, año de publicación, tipo de documento, clasificación, rúbrica (número de catálogo de novedades bibliográficas en el que está incluido), palabras del título, palabras de todo el documento, título de la revista (si se tratara de este tipo de documento) y fecha de entrada en la base de datos.

Es necesario recibir apoyo especializado para acceder a la base, así como para realizar las consultas ya que se requiere el conocimiento previo del programa informático.

Al establecer la conexión, aparecen en pantalla los datos del menú de apertura representados en las figuras 48 y 49:

INSTITUTO NACIONAL DE SERVICIOS SOCIALES

CONSULTA - MENU PRINCIPAL

BASES DE DATOS DISPONIBLES

- 1 - CATALOGACION
- 2 - ADQUISICIONES
- 3 - HIST. ADQUISICIONES
- 4 - PUBLICACIONES PERIODICAS

Elija la opción deseada :

V = Regresar al Menu anterior

CONEXION.SCR ECHO ROLL 1 NUML 13:48:51.20

INSTITUTO NACIONAL DE SERVICIOS SOCIALES

BASE DE DATOS : CATALOGACION

CONSULTA SENCILLA

Campos de Consulta :

- |                    |                       |                      |
|--------------------|-----------------------|----------------------|
| 1 -Referencia      | 7 -Desc. Geográficos  | 13 -Fuente           |
| 2 -Encabezamiento  | 8 -Lengua,Año,Tipodoc | 14 -Fecha de Entrada |
| 3 -Titulo          | 9 -Clasificación      |                      |
| 4 -Autores         | 10 -Rúbrica           |                      |
| 5 -Descriptoros    | 11 -Titulo (Palabras) |                      |
| 6 -Identificadores | 12 -Palabras Doc.     |                      |

No. CAMPO:

\$0

CM = Cons. Múltiple LA = Listar F. Aut. T = Listar Tesouro V = Volver M

CONEXION.SCR ECHO ROLL 1 NUML 13:49:23.16

Figura 48. Primeras pantallas del menú de la base de datos de la Biblioteca del Instituto de Migraciones y Servicios sociales

INSTITUTO NACIONAL DE SERVICIOS SOCIALES  
BASE DE DATOS : CATALOGACION CONSULTA COMPUESTA

\$1 176 respuesta(s) para: TIT=AS\*  
\$2 389 respuesta(s) para: TIT=TR\*

Elija la Opción deseada :

CS = Cons. Simple CM = Consulta Múltiple CH = Consulta Histórico  
H = Histórico IH = Inic. Histórico L = Listar LC Listar Campos Sec  
CONEXION.SCR ECHO ROLL 1 CAPS NUML 13:50:27.19

INSTITUTO NACIONAL DE SERVICIOS SOCIALES  
BASE DE DATOS : CATALOGACION CONSULTA COMPUESTA

\$1 176 respuesta(s) para: TIT=AS\*  
\$2 389 respuesta(s) para: TIT=TR\*  
\$3 565 respuesta(s) para: \$1 O \$2

\$ OPERADOR (Y-EXCP-O) \$

CS = Consul. Simple CM = Consul. Multiple V = Volver Menu Princ.  
CONEXION.SCR ECHO ROLL 1 CAPS NUML 13:51:06.19

Figura 49, Primeras pantallas del menú de la base de datos de la Biblioteca del Instituto de Migraciones y Servicios sociales

## Capítulo 6

# Evaluación del rendimiento de tesauros españoles en sistemas de recuperación de información

## **6.1. Introducción**

En teoría, si buscáramos exhaustivamente en toda la colección de documentos que compone una base de datos, llegaríamos a encontrar determinada información solicitada, pero con un desmesurado gasto de tiempo y esfuerzo. Para que la inversión de tiempo en dicha tarea se minimice, la información se indiza con objeto de señalar la existencia o no de documentos relevantes acerca de distintas materias, y se almacena siguiendo unos criterios de clasificación. Durante el proceso de indización se seleccionan los términos apropiados de un lenguaje documental para la representación del contenido del documento y se le asignan de acuerdo con las reglas establecidas. En el momento de la recuperación se siguen esos mismos criterios y, si las materias utilizadas para indizar y alimentar el sistema son coincidentes con las empleadas en el proceso de búsqueda, los documentos recuperados serán relevantes. De lo dicho se deduce la importancia que tienen en todo el proceso la calidad del lenguaje de clasificación o indización empleado, la calidad de la indización y, por último, la capacidad de seleccionar las materias adecuadas para formular una ecuación de búsqueda. Lógicamente, la persona que va a interrogar al sistema necesita conocer lo mejor posible la naturaleza del lenguaje documental utilizado para el proceso de normalización de la información, así como el grado de detalle con que se analizaron los documentos. En base a su conocimiento podrá formular preguntas con un nivel de profundidad o generalidad adecuados, aumentando la probabilidad de recuperar información relevante. Aún conociendo estas variables, raramente coinciden los descriptores, palabras clave, etc., utilizados en la indización de un documento y en una pregunta de recuperación, dándose una mayor tasa de coincidencia, como es natural, cuando se trata de cuestiones genéricas.

Como alternativa al lenguaje documental se puede recurrir al vocabulario empleado en el propio documento para realizar dicha representación; en tal caso, se hablará de indización y recuperación en lenguaje libre. Entre las ventajas que, a priori, se pueden reconocer al uso de un lenguaje controlado, está su capacidad para adaptarse con facilidad a necesidades de información más o menos específicas del usuario potencial. Con el test que describimos en este capítulo tratamos de contrastar sus ventajas y desventajas, por otra parte bien conocidas, en comparación con el uso del lenguaje no controlado, en SRI españoles, tratando con necesidades de información reales.

Como decíamos anteriormente, existen muchos elementos que influyen en la efectividad del proceso de recuperación de información, pero compartimos con Blair la idea de que *el más importante aspecto [...] concierne a cómo están representados los documentos en un sistema. El más rápido ordenador o las más sofisticadas técnicas de búsqueda no pueden superar los problemas derivados de una pobre representación [de naturaleza lingüística] de los documentos*<sup>1</sup>.

En los epígrafes que componen este capítulo tratamos los aspectos relativos a antecedentes, metodología, desarrollo de las búsquedas, criterios de evaluación, planteamiento de las estrategias y resultados obtenidos. El desarrollo del penúltimo apartado es de considerable extensión porque comprende las 120 ecuaciones de búsqueda realizadas tanto en lenguaje libre como con tesoro. Al final de la presentación de las ecuaciones de cada base de datos se comentan las incidencias que se produjeron en su planteamiento o desarrollo y se presenta un cuadro comparativo de los resultados obtenidos a partir de las estrategias. En el cuadro se registran el número de documentos relevantes recuperados con cada ecuación, así como las tasas de precisión y exhaustividad

---

<sup>1</sup>BLAIR, D.C. **Language and representation in information retrieval**. Amsterdam, etc.: Elsevier, 1990, p. 155.

conseguidas en lenguaje libre y con ayuda del tesoro. Finalmente se incluyen diagramas de barras que permiten comparar las tasas de precisión y exhaustividad en ambos tipos de lenguajes.

## 6.2. Antecedentes

Desde los años cincuenta se han venido realizando evaluaciones de los lenguajes utilizados en los Sistemas de Recuperación de Información (SRI) para la indización y la recuperación de documentos. Las características de las mismas, su nivel de profundidad y aspectos analizados, varían de unas a otras, pero todas, en una u otra forma, han alimentado el continuo debate entre los profesionales de la documentación acerca de la conveniencia y eficacia de los lenguajes controlados e incontrolados para realizar dichas operaciones.

Existe abundante literatura sobre la evaluación de estos sistemas. El primer test conocido fue el ASTIA-Uniterm test, realizado en 1953. La investigación se centraba básicamente en el análisis de los resultados, aunque no se tiene constancia del desarrollo del proyecto, a excepción de un breve trabajo publicado por GULL<sup>2</sup>. Para dicho trabajo se indizaron 15.000 documentos por dos grupos diferentes de especialistas en la materia, que provenían de la Armed Services Technical Information Agency (ASTIA) y de Documentation, Inc. El primer grupo utilizó la Lista de encabezamientos de materia creada por ASTIA, el segundo, el sistema uniterm. Se seleccionaron 98 demandas de entre las que había recibido ASTIA en el curso de sus actividades. Del análisis de los

---

<sup>2</sup>GULL, D. *Seven years of work on the organisation of materials in the special library*. *American Documentation*, v. 7 (1956) pp. 320-329.

resultados derivó una mayor capacidad y calidad de recuperación con el procedimiento de unitérminos que con el de encabezamientos de materia.

En 1954, Cleverdon y Thorne<sup>3</sup> realizaron un test sobre el sistema uniterm, cuyos resultados fueron inconclusos, pero permitieron sentar las bases metodológicas de futuras investigaciones.

Algunos estudios realizados sobre evaluación de SRI a lo largo de los últimos años evaluaron solamente los lenguajes utilizados para indizar y recuperar los documentos; otros se centraron en el proceso mismo de indización, etc. La gran mayoría se trataba de estudios de laboratorio, minuciosamente preparados con objeto de controlar cualquier posible variable, pero a partir de los años 80 se manifestó un creciente interés por analizar necesidades de usuarios reales. Los experimentos de Blair y Maron<sup>4</sup> (1984), así como los de Saracevic, Kantor, Chamis y Trivison<sup>5</sup> (1988) son indicativos de esa tendencia. No es posible mencionar aquí todos los tests llevados a cabo hasta la fecha, pero nos parecen de obligada referencia los proyectos Cranfield I y Cranfield II, cuya metodología y resultados han servido de guía a investigaciones posteriores, incluida la que realizamos en esta tesis doctoral.

---

<sup>3</sup>CLEVERDON, C.W. & THORNE, R.G. **A Brief Experiment with the Uniterm System of Coordinate Indexing for the Cataloging of Structural Data.** RAE Library Memorandum N° 7, AD 35004. Farnborough, England: Royal Aircraft Establishment, 1954.

<sup>4</sup>BLAIR, D & MARON, M.E. An evaluation of retrieval effectiveness for a full-text document retrieval system. **Commun. ACM**, 28, 3 (1984) pp. 281-299.

<sup>5</sup>SARACEVIC, T. [et al]. A study of information seeking and retrieving. I. Background and methodology; II. Users, questions and effectiveness; III. Searchers, searches and overlap. **Journal of the American Society for Information Science**, 39, 3 (1988) pp. 161-216.

El desarrollo del proyecto **Cranfield I**, a cargo de Cleverdon<sup>6</sup>, tuvo lugar en Gran Bretaña entre los años 1957 y 1962, auspiciado por la Association of Special Libraries and Information Bureaux (ASLIB), con la colaboración de la National Science Foundation (USA).

Se comparó el rendimiento que ofrecían cuatro sistemas de indización: se comprobaron la Clasificación Decimal Universal, un índice alfabético de materias, un sistema de clasificación facetado y el sistema uniterm. Se indizaron 18.000 documentos sobre ingeniería aeronáutica, utilizando los cuatro lenguajes de indización mencionados. Los documentos se trataron en grupos de 100 contabilizándose el tiempo de indización (entre 2 y 16 minutos). Se formularon 1.200 cuestiones en base a *documentos fuente*<sup>7</sup>, con las que se realizaron las búsquedas. Si el *documento fuente* era identificado, la búsqueda se juzgaba un éxito. El proceso se repitió tres veces para asegurar un mayor índice de imparcialidad en la operación.

En la segunda vuelta intervinieron las mismas personas que en la primera, con la incorporación de un ayudante de la College of Aeronautics Library y cinco estudiantes postgraduados. El tipo de preguntas formuladas fueron las mismas realizadas en la primera vuelta. El propósito de hacer intervenir a estas personas en el proceso de recuperación fue el de descubrir si algún sistema en particular resultaba más o menos satisfactorio cuando era empleado por usuarios que conocían el área temática pero que

---

<sup>6</sup>CLEVERDON, C.W. **Aslib Cranfield Research Project: report on the testing and analysis of an investigation into the comparative efficiency of indexing systems**. Cranfield: College of Aeronautics, 1962.

<sup>7</sup>En respuesta a cada cuestión había un único documento relevante (documento fuente), que era el utilizado para formular la pregunta. El uso del documento fuente para la preparación de las búsquedas ha sido, precisamente, el principal motivo de crítica del proyecto, porque este hecho podría haber influido en la probabilidad de recuperación de esos documentos. Swanson, D.R., por ejemplo, encabezó las críticas al procedimiento en el artículo: Evidence underlying the Cranfield results, publicado en **Library quarterly**, nº 35 (1965) pp. 1-20.

tenían poca experiencia en materia de recuperación de información<sup>8</sup>. A los estudiantes se les dió una mínima guía, consistente en la mera explicación de las normas de la operación. Al concluir la vuelta se analizaron los casos en los que se habían producido fallos para cambiar el procedimiento a seguir en la tercera vuelta de la investigación. Los resultados que habían sido positivos en la segunda vuelta se consideraron adecuados y no pasaron a analizarse en la tercera. Se trató de averiguar si la causa de los errores se debió a las dificultades de la pregunta formulada, a una indización inadecuada, a la búsqueda o al lenguaje documental utilizado.

El análisis de los fallos mostró que hubo muchos casos en los que la búsqueda había tenido éxito al utilizar un determinado lenguaje documental, pero había fallado en otros. En tales casos la razón del fracaso no tuvo nada que ver con el lenguaje empleado ni con la indización, sino que se debió por entero a que la persona que realizó la búsqueda a través de dicho lenguaje de indización no había conseguido construir una estrategia adecuada.

Los resultados se midieron sobre la base de dos parámetros: *exhaustividad* y *precisión*, que habían sido expresados algunos años antes por Fairthorne en las frases: *ABNO: All-But-not-Only* (alta exhaustividad) y *OBNA: Only-But-Not-All*<sup>9</sup> (alta precisión) que, a su vez, habían sido sugeridos por Perry, Kent y Berry<sup>10</sup> dos años antes. Fue Cleverdon, sin embargo, el primero en poner en práctica la evaluación de los lenguajes

---

<sup>8</sup>El proyecto Cranfield I demostró el hecho de que para un bibliotecario experimentado ni la falta de familiaridad con el sistema de indización y recuperación, ni el relativo desconocimiento de una determinada materia, presentan serios problemas. Este descubrimiento rebatía la creencia de que recuperar información con un lenguaje documental poco conocido y en un área temática desconocida para el usuario es una tarea difícil de realizar.

<sup>9</sup>FAIRTHORNE, R. Automatic retrieval of recorded information. *Computer journal* (1958) pp. 36-41.

<sup>10</sup>PERRY, J.; KENT, A.; BERRY, M. *Machine literature searching*. New York, etc.: Interscience Publishers, 1956.

documentales utilizados para el tratamiento y recuperación en SRI, basándose fundamentalmente en estos valores.

La *exhaustividad* mide la capacidad del sistema para recuperar documentos útiles, mientras que la *precisión* mide la habilidad de rechazar material no relevante<sup>11</sup>. Teniendo en cuenta estos índices, la flexibilidad de un sistema se podrá valorar en función de su capacidad para adaptarse a las necesidades de exhaustividad y precisión de los usuarios.

La *precisión* se halla dividiendo el número de documentos relevantes recuperados entre el número de documentos recuperados. La *exhaustividad* se calcula dividiendo el número de documentos relevantes recuperados entre el número de documentos relevantes existentes en la colección.

Los resultados demostraron que los cuatro lenguajes conseguían prácticamente los mismos niveles de rendimiento a la hora de recuperar los documentos fuente. El porcentaje de recuperación fue el siguiente:

---

<sup>11</sup>Un documento se juzga relevante cuando responde a una solicitud de información. Relevancia se emplea, a menudo, como término sinónimo de pertinencia, y como tal es considerado por algunos autores. Sin embargo, son mayoría los especialistas que establecen diferencias entre ellos. Lancaster, Foskett y Kemp, por ejemplo, hablan de relevancia cuando se refieren a la valoración que realizan una o varias personas en relación con una determinada solicitud de información, encontrando coincidencia entre pregunta y respuesta. Pertinencia sería la valoración que hace el usuario de una respuesta dada por un SRI a una necesidad concreta de información formulada por él mismo. Puede, como es natural, existir coincidencia en cuanto a las estimaciones de los dos índices, pero existe un sutil matiz diferenciador entre ambos: un documento considerado pertinente, no necesariamente puede ser valorado como útil o relevante por una persona ajena a la formulación de la cuestión.

Unitérminos	82%
Encabezamientos de materia	81'5%
Clasificación Decimal Universal	75'6%
Clasificación facetada	73'8%

La diferencia es poca, como muestran las cifras, con excepción del sistema facetado, que dió más baja tasa de exhaustividad que los otros tres. El bajo rendimiento de este lenguaje se atribuyó a la rigidez del orden fijo de combinación de términos que le caracteriza. En otras palabras, Cranfield no descubrió diferencias significativas en la eficacia de los lenguajes de indización estudiados, un resultado que ha sido reiteradamente contrastado en estudios comparativos posteriores.

En cualquier caso, la importancia del primer proyecto Cranfield no estriba, en los resultados obtenidos, tanto como en el método utilizado, el primero en su género, que sirvió de modelo a los test realizados posteriormente.

**El segundo proyecto Cranfield (1963-1966)**, tenía como principal objetivo investigar los componentes de los lenguajes de indización y los efectos que esos diversos componentes tenían sobre el rendimiento de los SRI.

El experimento se desarrolló de muy distinta forma al primero. Para llevarlo a cabo se construyeron 33 tipos diferentes de lenguajes de indización variando terminologías y estructuras. Cada lenguaje variaba igualmente tanto en el uso de términos simples y compuestos como en la incorporación de jerarquías. En todos ellos se controlaron sinónimos y homógrafos.

El corpus lo formaban 1.400 documentos que trataban de aeronáutica. Se obtuvieron un total de 221 formulaciones de búsqueda preguntando a autores de una serie de documentos seleccionados para reconstruir las cuestiones que originalmente dieron lugar a la redacción o preparación de dichos trabajos. La búsqueda se hizo preparando estrategias basadas en diferentes niveles de coordinación de términos, y el rendimiento se midió por la recuperación de documentos previamente identificados como relevantes a la pregunta. En este test, por lo tanto, como señala Foskett<sup>12</sup>, la valoración de la relevancia se realizó *a priori*, examinando todos los documentos de la colección que se iban a analizar en relación con todas las cuestiones formuladas.

Asimismo, se tuvieron en cuenta los índices de exhaustividad y precisión, tratando de averiguar qué lenguaje de indización aumentaba o disminuía esos parámetros. Los resultados del proyecto se pueden resumir en un mejor rendimiento de los lenguajes formados por términos simples, de los que se habían eliminado los casos de sinonimia y todos aquellos términos que generaban ambigüedad. Otra conclusión que pudo extraerse del test fue que la simple coordinación de términos resultó ser el recurso más efectivo para aumentar la precisión, independientemente de que dichos términos formaran parte o no de un lenguaje controlado. Este resultado, que parecía echar por tierra la necesidad de utilizar los lenguajes documentales, fue ampliamente rebatido por el hecho de haberse dado en una experimentación artificial, en la que se mantenían bajo control variables que, en necesidades de información reales, son difícilmente controlables.

A la vista de los resultados se pudo mantener la hipótesis de que existía una relación inversa entre exhaustividad (número de documentos que encontramos cuando realizamos una búsqueda) y precisión (probabilidad de que correspondan a una necesidad

---

<sup>12</sup>FOSKETT, A.C. **The subject approach to information**. 5ª ed. London: Clive Bingley, 1996, p. 16.

de información). Dicha hipótesis explica que, a medida que se amplía el campo de búsqueda, se encuentra más cantidad de información.

La confirmación de esa proporción inversa implicaría que es inútil cualquier intento de maximizar ambos índices en un mismo sistema de información. Foskett sostiene, sin embargo, que *se pueden dar casos en los que, al examinar los resultados de búsquedas individuales, encontremos que se ha conseguido un 100% de exhaustividad con un 100% de precisión o, por el contrario, puede haber casos en los que ambos dan como resultado cero*<sup>13</sup>. Van Slype también considera posible, que ambos criterios mejoren simultáneamente en una misma consulta, siempre *se pongan en funcionamiento nuevos medios (cualificación de los documentalistas, calidad del lenguaje documental y sofisticación del programa informático de búsqueda)*<sup>14</sup>.

### 6.3. Metodología

En nuestro trabajo experimental, a diferencia de los mencionados proyectos, no analizaremos los resultados de la indización para introducir datos en el sistema con uno u otro lenguaje, sino que vamos a establecer la comparación entre la capacidad de recuperación de dos diferentes lenguajes: libre, basado fundamentalmente en términos simples o unitérminos, y los tesauros utilizados en los seis SRI consultados.

Teníamos que elegir entre realizar un experimento de evaluación artificial, controlando a priori todas las posibles variables: niveles de indización, tipología de los

---

<sup>13</sup>FOSKETT, A.C. Opus cit., p. 85.

<sup>14</sup>SLYPE, G. Van. **Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales**. Madrid: Fundación Germán Sánchez Ruipérez, 1991, p. 194.

documentos, cantidad y calidad de los documentos, recursos automáticos de búsqueda y lenguajes documentales, o bien, basar nuestra evaluación en necesidades de información reales, para lo cual contábamos con serias limitaciones. La primera, y fundamental, la constituía el hecho de no contar con un equipo de personas para realizar el análisis de un número representativo de documentos, dado que se trataba de un trabajo individual. Optamos por analizar el comportamiento de seis tesauros frente al lenguaje libre, en la recuperación de información en bases de datos reales. Las bases elegidas para este propósito también fueron seis, de las cuales tres son del área de las ciencias sociales y otras tres de ciencia y tecnología. Así pues, el objetivo de la investigación se concreta en el examen del rendimiento de tales tesauros, desde la perspectiva del usuario, en el contexto de las bases de datos en las que se utilizan para el tratamiento documental.

El escaso número de bases de datos con tesauro propio, unido a dificultades derivadas de la ausencia de automatización y a traslados de sede de algunos centros, en el momento de realizar este trabajo, hicieron difícil la elección de las bases que precisábamos para hacer el test. Tras comprobar las prestaciones de unas y otras, optamos por las seis, analizadas en el anterior capítulo, que, si bien no son totalmente homogéneas, reúnen las condiciones necesarias, es decir, todas ellas son bases de datos bibliográficas que cuentan con tesauro propio y son de acceso directo. El volumen de documentos con que cuentan difiere sustancialmente:

Biología animal	11.820
Electrotecnica	1.360
Medio ambiente	26.362
Centro Doc. Mujer	12.600
Psicología	18.089
Migraciones y S. sociales	21.392

pero consideramos que esta particularidad, en vez de ser un inconveniente, constituye un valor añadido para la experimentación, ya que permite valorar el rendimiento de tesauros destinados a tratar fondos documentales de diverso tamaño.

Para realizar el análisis comparativo preparamos, en principio, un cuestionario de posibles preguntas a formular en cada base de datos, para lo cual fue imprescindible la lectura de obras y revistas especializadas en las materias a consultar, dada su heterogeneidad y grado de especialización.

Teniendo en cuenta que el trabajo lo iba a realizar una sola persona, las preguntas se plantearon desde un primer momento a un nivel de amplitud suficiente como para poder conseguir resultados válidos para la investigación y, al propio tiempo, lo más específicos que fuera posible, teniendo en cuenta que para poder validar los resultados, iba a ser necesario leer cada una de las referencias derivadas de las búsquedas. En algunos casos, incluso, se hizo necesario consultar los documentos completos al no ser posible hacer una valoración de su relevancia con el simple análisis de título y descriptores (en aquellas bases en las que el registro no contiene campo de resumen). El conocimiento previo del nivel de indización media que realizan las distintas bases también nos sirvió para plantear estrategias adecuadas a nuestros fines.

Se procedió, en primer lugar, a formular una ecuación de búsqueda en lenguaje natural, realizándose una segunda estrategia en lenguaje controlado. Inmediatamente después, se evaluó la relevancia de cada documento recuperado con uno y otro tipo de lenguaje.

Cuando se recuperó en lenguaje libre, el sistema trató de localizar la información en todos los campos del registro (salvo los que se componen de palabras vacías o forman

parte de claves), por lo que tomó también los unitérminos que componen los descriptores. La búsqueda en esta modalidad se realizó, pues, consultando los campos de título, título original, descriptores principales y secundarios, topónimos e identificadores, resumen e índices, comunes todos ellos a la gran mayoría de las bases de datos. Las palabras clave compuestas se hallaron con un operador de adyacencia (para lo cual tenían que estar unidas en el texto). No todos los centros admitían el uso de palabras clave compuestas para la preparación de las ecuaciones, de modo que se hizo necesario recurrir a la *post-coordinación en algunas de ellas. Lógicamente las palabras clave utilizadas para recuperar con lenguaje natural no siempre fueron coincidentes con los descriptores componentes de los tesauros y, en ocasiones, lo fueron sólo parcialmente. En cuanto al campo de título, las palabras clave que conformaban las ecuaciones de búsqueda aparecían a veces en el mismo, si bien, la búsqueda en este campo presenta el problema de que se pierde la información existente en títulos de idioma extranjero, por lo que su eficacia es relativa y está condicionada a la lengua en la que está escrito el documento. El campo de resumen permite una mayor efectividad en la búsqueda en lenguaje libre, lo que, a su vez, produce, como indicábamos anteriormente, un detrimento en lo que a precisión informativa se refiere. Solamente consultamos el campo de índices de materia en la base de datos de Servicios sociales. Este recurso fue utilizado para tratar de suplir la ausencia del campo de resumen en dicha base. Si bien es cierto que el control ejercido sobre el vocabulario en los índices de las materias contenidas en los documentos es mucho mayor que el posible control existente en un resumen, consideramos que su consulta potenciaba los resultados conseguidos con los otros campos de búsqueda y, por tanto, permitía la homogeneidad con las demás bases.*

La recuperación en lenguaje controlado se realizó consultando los campos de descriptores. Los tesauros que cubren las bases sugieren términos relevantes para la búsqueda, de modo que la preparación de las ecuaciones con este procedimiento resultó



más sencillo; no hacía necesario pensar en posibles términos sinónimos o polisémicos para delimitar los conceptos. La última fase del proceso consistió en imprimir o en exportar a disquete cada una de las referencias resultantes para proceder a su posterior análisis.

El procedimiento seguido para la valoración de los resultados fue el propuesto por Cleverdon, teniendo en cuenta los índices de precisión y exhaustividad. En respuesta a cada cuestión se recuperaron documentos con un índice de relevancia diverso; algunos fueron totalmente relevantes, otros respondían de forma dudosa a la pregunta formulada, y otros tenían una correspondencia marginal con dicha pregunta. Para el análisis se tomaron sólo en cuenta aquellos totalmente relevantes. Como nuestro grado de conocimiento de las diferentes materias no era homogéneo, solicitamos el consejo de especialistas en el contenido temático de las distintas bases de datos cuando nos surgieron dudas al establecer los criterios de relevancia, de tal manera que al nuestro se sumara el del potencial usuario conocedor de la materia. En cualquier caso, tratamos de valorar los documentos con la mayor objetividad posible, conocedores de que la medida de la relevancia es una estimación subjetiva que depende del nivel de conocimientos de cada individuo. *El problema -escribe Foskett al respecto -reside en el hecho de que los lectores buscan información partiendo de su nivel de conocimiento -su marco de referencia- con el mínimo esfuerzo, mientras que los autores presentan la información bajo su propio marco de referencia; cada uno de nosotros tiene su marco propio, por lo que la coincidencia nunca puede ser exacta. Podemos diseñar nuestros sistemas de recuperación de información para optimizar la probabilidad de coincidencia entre las preguntas de los usuarios con las respuestas que consiguen, pero aceptando de hecho que nunca será perfecta*<sup>15</sup>.

---

<sup>15</sup>FOSKETT, A.C. Op. cit., p. 15.

#### 6.4. Desarrollo de las búsquedas

Antes de comenzar cada consulta fue necesario aprender el lenguaje específico de cada sistema. Si bien es cierto que a pesar de las diferencias entre unos y otros, tienen todos rasgos comunes que permiten adaptarse fácilmente a su manejo, ello implica un esfuerzo más a realizar en el proceso. A este respecto escribe Amat: *Un inconveniente en relación con la figura del distribuidor de bases de datos radica en el hecho de que cada uno tiene que desarrollar su propio programa de recuperación de información o lenguaje de interrogación. Este hecho provoca que no se pueda trabajar con un único lenguaje de interrogación, sino que se han de conocer los diferentes lenguajes de los distribuidores con los que se trabaja*<sup>16</sup>.

El manejo de las bases en CD-Rom fue muy sencillo, dándonos la posibilidad de trabajar con autonomía. En los SRI en los que la consulta se hizo en línea, fue necesaria la ayuda del documentalista o especialista en materia de recuperación; no obstante, estuvimos presentes durante el desarrollo de todas las búsquedas, en una u otra modalidad, puesto que se hacía necesaria la modificación de las estrategias con frecuencia. A veces, la combinación de conceptos daba como resultado una ausencia total de referencias, obligándonos a renunciar a los aspectos menos importantes de la cuestión, o a ampliar el alcance de algún concepto.

Para la preparación de las ecuaciones se analizó el área temática de las preguntas en detalle, de forma que para construir cada ecuación se tuvieran en cuenta los conceptos y relaciones que pudieran sugerir alternativas flexibles para preparar la búsqueda utilizando la capacidad de interacción de los SRI. No sólo se desarrollaron las estrategias

---

<sup>16</sup>AMAT, N. *La documentación y sus tecnologías*. Madrid: Pirámide, 1994, p.139.

iniciales sino que se fueron evaluando las referencias encontradas. Seguimos con este procedimiento la técnica sugerida por Van Rijsbergen<sup>17</sup> y Croft<sup>18</sup> consistente en partir de un documento considerado a priori relevante en relación con la pregunta formulada y tratar de localizar otros similares basándonos en él.

Nuestro desconocimiento de las materias tratadas en algunas bases de datos también nos hizo recurrir a la alternativa denominada *pearl growing* para preparar estrategias, que *consiste en analizar un documento (completo o título y resumen) conocido como representativo o de importancia (relevancia) en el área de conocimiento deseada y, en base al vocabulario que emplea, preparar la estrategia*<sup>19</sup>. Estas alternativas fueron posibles gracias a la naturaleza heurística de la recuperación en línea, que nos permitió ir modificando las estrategias a la vista de los resultados que íbamos obteniendo. En muchos casos, el examen de las referencias recuperadas nos mostró nuevos términos que pudimos incluir en nuevas ecuaciones.

A pesar de que utilizamos estos recursos, que nos fueron de gran utilidad, al tratar con algunos temas era difícil darnos cuenta de cualquier variación posible en la denominación de los conceptos presentes en las preguntas, de modo que para resolver este inconveniente solicitamos la ayuda de expertos en las materias. En el campo de las ciencias sociales fueron más frecuentes las dudas que en el área de ciencia y tecnología. En aquel campo se dieron en ocasiones falsas interpretaciones de las preguntas y surgieron dificultades para combinar términos; este último inconveniente se resolvió, a

---

<sup>17</sup>Van RIJSBERGEN, C.J. **Information retrieval**. 2ª ed. London: Butterword & Co., 1979.

<sup>18</sup>CROFT, W.B. A model of cluster searching based on classification. **Information systems**, nº 5 (1980) pp. 189-195.

<sup>19</sup>HARTLEY, R.J. **Online searching: principles & practice**. London, [etc]: Bowker-Saur, 1990, p. 171.

menudo, con el uso de palabras clave o descriptores precoordinados. En algunos supuestos, la inclusión de los conceptos de la pregunta conectados con el operador [Y] dió un resultado preciso de la información requerida. En ciertas consultas se amplió la búsqueda simplemente omitiendo uno de los conceptos de partida.

Nos queda, por último señalar que, en ningún caso se limitaron los registros a determinado tipo de documentos, sino que se tuvieron en consideración las referencias derivadas de monografías, artículos, actas de congresos, etc... Tampoco se tuvo en cuenta la fecha de publicación (antigüedad) de los documentos, dada la heterogeneidad de las bases. Se trató de encontrar, en todo momento, toda la información disponible sobre la materia, no sólo las referencias más pertinentes, para ello se formularon las cuestiones, por lo común, partiendo de lo general a lo particular, hasta llegar al más reducido o específico nivel posible sin correr riesgos de perder información relevante.

#### **6.4.1. Aspectos y consideraciones para la preparación de las estrategias de búsqueda**

A continuación describiremos aquellos aspectos que tuvimos que considerar en la preparación de las estrategias de búsqueda. Estos elementos hacen referencia al tipo de términos utilizados, niveles de concreción, problemas de sinonimia y polisemia, operadores utilizados, problemática en el uso del lenguaje libre y otras variables que hubimos de tener en cuenta a fin de diseñar una óptima estrategia de búsqueda. Dichos aspectos pueden ser señalados en los siguientes puntos:

1) Se analizaron los temas generales concernientes a cada pregunta en sus facetas constituyentes, así como los términos sinónimos y cuasisinónimos de cada materia. Se ordenaron y agruparon en conjuntos y mediante el álgebra de Boole - utilizada en todas

las bases -, se precisaron los temas hasta conseguir las referencias que respondían al perfil de búsqueda solicitado. Se visualizaron en pantalla las muestras obtenidas a fin de comprobar que los documentos se centraban en la materia deseada. Según la respuesta obtenida, se ampliaba o reducía la búsqueda mediante diversos recursos.

2) Con objeto de conseguir en la búsqueda inicial un alto nivel de exhaustividad, se incluyó más de una palabra para describir cada concepto. Si el número de registros resultante indicaba que alguno de los términos empleados resultaba demasiado general, se procedía a realizar una nueva combinación de términos, omitiendo alguno de los más generales, obteniendo así un nivel adecuado de referencias.

3) Algunas bases de datos ofrecían la posibilidad de utilizar términos genéricos incluidos en los tesauros que facilitaban la amplitud de las búsquedas. Su empleo supuso un significativo ahorro de tiempo y esfuerzo, puesto que hizo innecesaria la localización de terminología precisa, al tiempo que simplificó la construcción de la estrategia. Al elegir estos descriptores genéricos se evitó el uso de palabras que describían conceptos demasiado imprecisos que podían generar un alto grado de ambigüedad, tales como *sistema, medida, problema, etc.*,

4) A fin de alcanzar un alto nivel de precisión, se recurrió al uso de términos específicos y, en su caso, a la combinación de los mismos. Como es lógico, los términos específicos tienen la particularidad de concentrarse en unos pocos documentos de la colección, lo que facilita su identificación. Así, por ejemplo, en la base de datos de Electrotecnia, el alto nivel de precisión concretó los términos utilizados en las estrategias en unos pocos documentos de la colección. Si tenemos en cuenta el limitado número de documentos que la componen, es razonable que en algunos casos las tasas de precisión y exhaustividad resultantes se aproximasen y que incluso fueran coincidentes.

5) En algunas consultas ocurrió que las preguntas incluían una gran variedad de conceptos diferentes, por lo que se trató de representarlos por medio de términos unidos por el operador [Y]. En tales casos se conseguía un alto nivel de precisión y, en ocasiones, dio como resultado el silencio. Este fue el caso de la tercera consulta de la base de datos de Servicios Sociales, realizada en lenguaje controlado.

6) El operador de negación [NO], que permite combinar expresiones alternativas y agrupar términos representativos de un mismo concepto, restringiendo así el volumen de información recuperada, fue un recurso empleado en contadas ocasiones, ya que, aunque facilita las búsquedas en algunos casos, su uso implica el riesgo de perder información relevante. En su lugar, optamos por recurrir a los términos genéricos y relacionados de los tesauros o por utilizar expresiones alternativas en lenguaje libre.

7) Como consecuencia del uso de términos ambiguos, imprecisos o inapropiados, en ocasiones se dieron falsas combinaciones, dando como resultado la recuperación de referencias que no eran lo bastante precisas. Para eliminar estos inconvenientes se consultaron los términos de indización de alguna referencia relevante recuperada con anterioridad, lo que nos permitió contextualizar los términos de la pregunta y replantear la estrategia.

8) En algún caso se dió la circunstancia de que el lenguaje controlado no disponía de términos adecuados para describir los conceptos requeridos. Dado que nuestro estudio consistía en establecer una comparación entre el lenguaje libre y el controlado, optamos por incorporar términos más genéricos para que el tesoro pudiera dar una respuesta satisfactoria. Sin embargo, y como cabía esperar, el resultado fue impreciso, ya que se alejaba de las necesidades de información que se precisaban en la pregunta.

9) Por último, cabe decir que cuando se utilizó el lenguaje libre se tuvieron en cuenta todas las posibles variaciones de un mismo concepto, así como el truncado de prefijo y sufijo. Asimismo, en su caso se incluyeron las variaciones en singular y plural de cada palabra.

### 6.5. Criterios de evaluación

La evaluación de la cantidad y calidad de la información recuperada, se hizo teniendo en cuenta las tasas de efectividad, es decir, *precisión* y *exhaustividad*. Los resultados se presentan como porcentajes de dichos valores. Se consideraron relevantes las respuestas que correspondían adecuadamente al objetivo de la búsqueda.

En la práctica es difícil medir estos parámetros, debido a que la estimación de la relevancia es subjetiva. Hallar el índice de precisión genera pocos problemas, excepto cuando en una determinada búsqueda no se recupera ningún documento. La exhaustividad presenta muchas complicaciones, incluso cuando se trata de pequeñas colecciones ya que hallar este índice, que es inevitablemente un valor relativo, requiere que cada documento de la colección sea contrastado en relación con cada consulta sobre un tema determinado, ésto es, exige el conocimiento del número total de documentos relevantes de la colección con respecto a la pregunta. Ante la imposibilidad de determinar el volumen de documentos que tiene una base de datos sobre un tema en particular, diferentes especialistas han propuesto métodos alternativos que tratan de salvar la dificultad para calcular este valor.

Lancaster<sup>20</sup> utiliza en dos ocasiones un método de muestreo para estimar el índice de exhaustividad de una gran base de datos, es decir, midiendo la relevancia de un conjunto de documentos de la colección. Salton propone el mismo procedimiento y sostiene que la exhaustividad no es un valor exacto, sino una estimación del número total de documentos relevantes de la colección: *La valoración de la relevancia se hace en base a un subconjunto de documentos de la colección. Alternativamente, una consulta dada puede ser procesada por una variedad de diferentes métodos de búsqueda y recuperación, dando por supuesto que todos los documentos relevantes van a ser recuperados por medio de dichas búsquedas. Los resultados se combinan entonces en una única lista de resultados. La lista de documentos relevantes se obtiene mediante la valoración de la relevancia de esa lista de resultados*<sup>21</sup>. Van Slype considera dos posibilidades para determinar la exhaustividad: *ya sea sistemáticamente, examinando las referencias una a una (lo que tiene el riesgo de durar mucho); o bien interrogando de nuevo el fondo con una serie de ecuaciones muy amplias (con pocos o ningún Y, con muchos O), incluso basándose en la clasificación*<sup>22</sup>.

En el test que realizamos, hallamos las tasas de exhaustividad combinando las técnicas de muestreo, las de clasificación y la interrogación al sistema por los términos más genéricos de cada pregunta. Entre los documentos recuperados por medio de esta última alternativa, se encontraron muchos no relevantes, lógicamente, es decir, se dió un altísimo índice de ruido documental. Se imprimieron, por ejemplo, todos los documentos

---

<sup>20</sup>LANCASTER, F.W. *Evaluation of the Medlars demand search service*. Bethesda, Md.: National Library of Medicine, 1968; e *Information retrieval systems: characteristics, testing and evaluation*. 2ª ed. New York: Wiley, 1979.

<sup>21</sup>SALTON, G. & MCGILL, M.J. *Introduction to modern information retrieval*. New York, [etc]: McGraw-Hill Publishing Company, 1983, pp.166-167.

<sup>22</sup>SLYPE, G.v. Op. cit., p. 193.

en los que figurara el descriptor o palabra clave *insectos*, en la base de datos de Biología animal. En la estrategia realizada para hallar la tasa de precisión, el número de documentos que respondían a la cuestión: *Fisiología de la reproducción en los insectos*, fue de tan sólo 4, pero, al buscar por el término general *insectos*, tuvimos que analizar un total de 1.558 documentos. La proporción en este caso es exagerada, pero da una idea del propósito que perseguíamos y que se ha conseguido: no excluir del análisis ninguna información existente en la colección sobre cada tema.

El volumen de documentos extraídos mediante el último método mencionado nos hizo desistir, en un principio, de la tarea -en total se recuperaron 11.906 referencias-. Decidimos, finalmente, seguir adelante convencidos de que, aunque minucioso, era el único procedimiento que garantizaba resultados fiables. Tras el largísimo proceso de análisis teníamos la certeza de haber dado con toda la información relevante de las bases en relación con los temas propuestos.

Para hallar la exhaustividad mediante técnicas de muestreo, se procedió de la siguiente forma: se seleccionó en primer lugar la muestra generando un listado de números aleatorios por medio del generador de números aleatorios del compilador Pascal, cambiando la semilla para cada base de datos. Se hizo, a continuación, una estimación del tamaño de la muestra:

Para estimar el porcentaje de documentos relevantes definimos  $A_i$ , variable dicotómica, de forma que:

$$A_i < \begin{cases} 0 & \text{si el registro } i \text{ no es relevante} \\ 1 & \text{si el registro } i \text{ es relevante} \end{cases}$$

Por lo que el estimador del porcentaje será:

$$\hat{p} = \frac{\sum A_i}{n}$$

$\Sigma$  = suma

$\hat{p}$  = porcentaje estimado de registros relevantes  
 $n$  = número de registros de la muestra

Como se iba a realizar un muestreo aleatorio simple, sin reposición, la varianza del estimador de  $[p]$  sería:

$$V(\hat{p}) = \frac{N-n}{N-1} \cdot p q$$

$N$  = tamaño de la población  
 $n$  = tamaño de la muestra  
 $p$  = probabilidad de que  $A_i$  valga 1  
 $q$  = probabilidad de que  $A_i$  valga 0

Por lo tanto el error estimado será:

$$E = k^2 \sqrt{\frac{N \cdot npq}{N \cdot 1n}}$$

Según Kish, *la definición básica del error estándar de la media es igual para cualquier diseño de muestra: es el error estándar de la distribución de muestreo para ese determinado diseño de muestra*<sup>23</sup>.

---

<sup>23</sup>KISH, L. Selección de la muestra. En: *Los métodos de investigación en las Ciencias sociales*. Barcelona: Paidós, 1953, reimp. 1987, p. 177.

Para obtener un tamaño de muestra significativa al 95% y con un error no superior al 0.02 (diferencia entre el valor estimado y el valor real), despejamos [n]:

$$n = \frac{K^2 N p q}{E^2 (N-1) + K^2 p q}$$

Como desconocemos la varianza la sustituimos por 0.5 (1-0.5) = 0.25, que es el valor máximo que podría tomar.

El tamaño de la muestra de las cinco bases sometidas a este procedimiento fue el siguiente:

Biología animal	1.995 + 100 = 2.095
Medio ambiente	2.200 + 100 = 2.300
Centro Doc. Mujer	2.016 + 100 = 2.116
Psicología	2.119 + 100 = 2.219
Servicios sociales	2.158 + 100 = 2.258

La base de datos de Electrotecnia se analizó en su totalidad, dado el reducido volumen de sus fondos, por lo que no fue necesario proceder mediante este método para conocer la colección.

En todas las bases, como puede observarse, se incluyeron 100 documentos de reserva, en previsión de posibles bajas que hubieran sufrido las colecciones, o a cualquier otra circunstancia que alterara el volumen de la muestra.

## 6.6. Desarrollo de las estrategias

### 6.6.1. Base de datos del CSIC de Biología animal

#### Consulta 1. *Migración de aves en España*

##### Búsqueda en lenguaje libre

0 [migración]	157
1 [aves]	890
2 [España]	79262
3 [emigración]	2
4 [inmigración]	6
5 [.L0 .O .L3 .O .L4]	159
6 [.L1 .Y .L2 .Y .L5]	26

$$\text{precisión} = \frac{26}{26} = 1$$

$$\text{exhaustividad} = \frac{26}{41} = 0'63$$

##### Búsqueda en lenguaje controlado

0 [migración]	135
1 [aves migratorias]	46
2 [aves]	821
3 [.L1 .O .L2]	821
4 [España]	79262
5 [.L0 .Y .L3 .Y .L4]	25

$$\text{precisión} = \frac{25}{25} = 1$$

$$\text{exhaustividad} = \frac{25}{41} = 0'60$$

**Consulta 2. Composición bioquímica de huevos de moluscos**

Búsqueda en lenguaje libre

0 [gastrópodo*]	345
1 [mollus*]	921
2 [lípidos .O carbohidratos .O proteínas*]	929
3 [embrio*]	618
4 [huevo*]	230
5 [(.L0 .O .L1) .Y .L2 .Y (.L3 .O .L4)]	2

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{4} = 0'50$$

Búsqueda en lenguaje controlado

0 [mollusca]	918
1 [huevos]	214
2 [embriones]	152
3 [ .L0 .Y (.L1 .O .L2)]	4

$$\text{precisión} = \frac{4}{4} = 1$$

$$\text{exhaustividad} = \frac{4}{4} = 1$$

### Consulta 3. Fauna marina gallega

#### Búsqueda en lenguaje libre

0 [fauna]	560
1 [mar*]	15575
2 [Galicia .O galleg*]	2260
3 [.L0 .Y .L1 .Y .L2]	9

$$\text{precisión} = \frac{5}{9} = 0'55$$

$$\text{exhaustividad} = \frac{5}{5} = 1$$

#### Búsqueda en lenguaje controlado

0 [fauna marina]	43
1 [Galicia]	1639
2 [.L0 .Y .L1]	5

$$\text{precisión} = \frac{3}{5} = 0'60$$

$$\text{exhaustividad} = \frac{3}{5} = 0'60$$

**Consulta 4. Diferenciación celular en el desarrollo de embriones de anfibios.**

**Búsqueda en lenguaje libre**

0 [diferenciación]	239
1 [celul*]	1162
2 [anfibio*]	88
3 [embrion*]	466
4 [.L0 .Y .L1 .Y .L2 .Y .L3]	2

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{8} = 0'25$$

**Búsqueda en lenguaje controlado**

0 [citodiferenciación]	95
1 [amphibia]	325
2 [desarrollo embrionario]	330
3 [.L0 .Y .L1 .Y .L2]	3

$$\text{precisión} = \frac{3}{3} = 1$$

$$\text{exhaustividad} = \frac{3}{8} = 0'37$$

### Consulta 5. **Ecología de sistemas forestales**

#### Búsqueda en lenguaje libre

0 [ecología]	2064
1 [bosques]	247
2 [sistemas forestales]	3
3 [ .L1 .O .L2]	250
4 [ .L0 .Y .L3]	55

$$\text{precisión} = \frac{19}{55} = 0'34$$

$$\text{exhaustividad} = \frac{19}{29} = 0'65$$

#### Búsqueda en lenguaje controlado

0 [ecología]	1344
1 [ecosistemas forestales]	90
3 [ .L0 .Y .L1]	13

$$\text{precisión} = \frac{13}{13} = 1$$

$$\text{exhaustividad} = \frac{13}{29} = 0'44$$

**Consulta 6. Fósiles vertebrados del pleistoceno en España**

**Búsqueda en lenguaje libre**

0 [fósiles]	1242
1 [vertebrados]	509
2 [pleistoceno]	199
3 [España]	79262
4 [.L0 .Y .L1 .Y .L2 .Y .L3]	74

$$\text{precisión} = \frac{74}{74} = 1$$

$$\text{exhaustividad} = \frac{74}{101} = 0'73$$

**Búsqueda en lenguaje controlado**

0 [fósiles vertebrados]	471
1 [pleistoceno]	193
2 [España]	79262
3 [.L0 .Y .L1]	77
4 [.L2 .Y .L3]	71

$$\text{precisión} = \frac{71}{71} = 1$$

$$\text{exhaustividad} = \frac{71}{101} = 0'70$$

### Consulta 7. Fisiología de la reproducción en los insectos

#### Búsqueda en lenguaje libre

0 [insect*]	2038
1 [reproduc*]	1158
2 [fisiolog*]	1414
3 [.L0 .Y .L1]	30
4 [.L2 .Y .L3]	4

$$\text{precisión} = \frac{4}{4} = 1$$

$$\text{exhaustividad} = \frac{4}{13} = 0'30$$

#### Búsqueda en lenguaje controlado

0 [insecta]	1546
1 [reproducción]	979
2 [fisiología]	1188
3 [.L0 .Y .L1]	20
4 [.L2 .Y .L3]	4

$$\text{precisión} = \frac{4}{4} = 1$$

$$\text{exhaustividad} = \frac{4}{13} = 0'30$$

### Consulta 8. Cultivo de la dorada

#### Búsqueda en lenguaje libre

0 [dorada]	27
1 [cultivo]	1692
2 [acuicultura]	225
3 [.L1 .O .L2]	1868
4 [.L0 .Y .L3]	8

$$\text{precisión} = \frac{7}{8} = 0'87$$

$$\text{exhaustividad} = \frac{8}{13} = 0'61$$

#### Búsqueda en lenguaje controlado

0 [dorada]	18
1 [sparus aurata]	16
2 [acuicultura]	218
3 [piscicultura]	79
4 [.L0 .O .L1]	18
4 [.L2 .O .L3]	291
5 [.L4 .Y .L5]	11

$$\text{precisión} = \frac{11}{11} = 1$$

$$\text{exhaustividad} = \frac{11}{13} = 0'84$$

**Consulta 9. Anidamiento de aves acuáticas**

**Búsqueda en lenguaje libre**

0 [anidamiento]	1
1 [nidos]	35
2 [aves acuáticas]	113
3 [.L0 .O .L1]	36
4 [.L2 .Y .L3]	1

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{11} = 0'09$$

**Búsqueda en lenguaje controlado**

0 [nidificación]	114
1 [aves acuáticas]	113
2 [.L0 .Y .L1]	9

$$\text{precisión} = \frac{9}{9} = 1$$

$$\text{exhaustividad} = \frac{9}{11} = 0'81$$

**Consulta 10. Distribución geográfica de micromamíferos**

**Búsqueda en lenguaje libre**

0 [distribución]	3351
1 [geografi*]	2406
2 [micromamíferos]	58
3 [España]	79262
4 [.L0 .Y .L1 .Y .L2 .Y .L3]	2

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{3} = 0'66$$

**Búsqueda en lenguaje controlado**

0 [biogeografía]	1768
1 [micromammalia]	88
2 [España]	79262
3 [.L0 .Y .L1 .Y .L2]	2

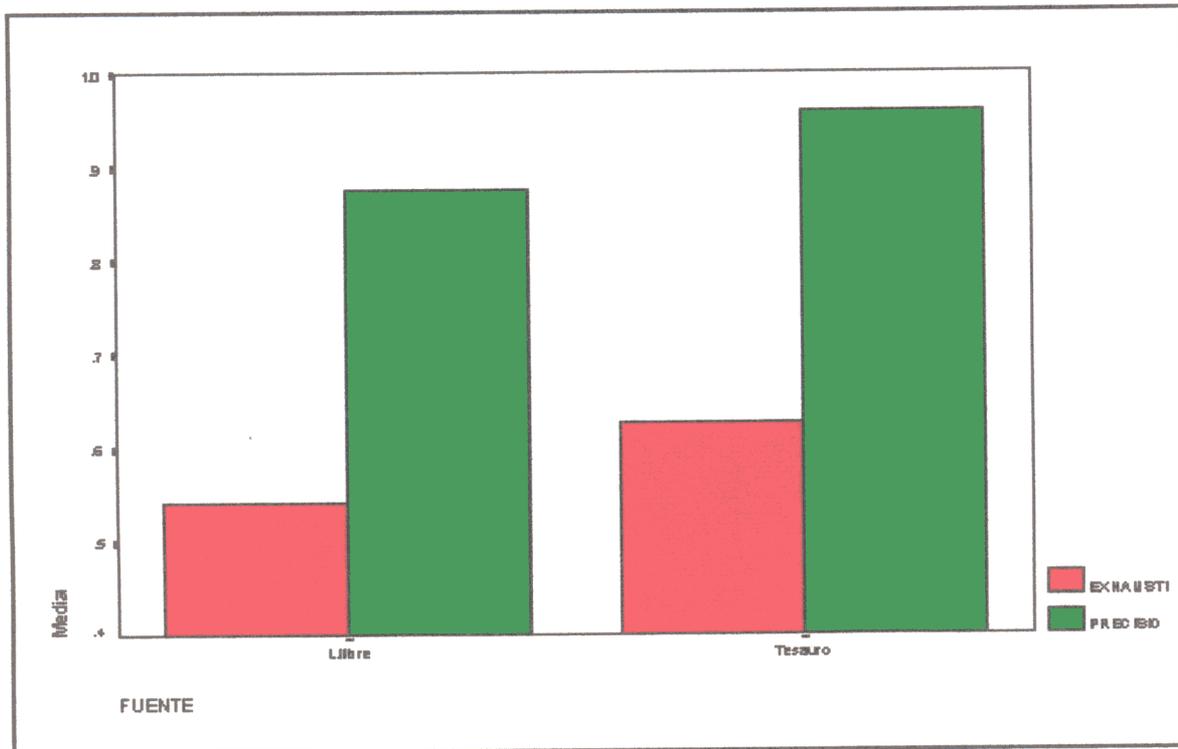
$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{3} = 0'66$$

**Cuadro comparativo de resultados:**

N° de Búsqueda	Lenguaje libre			Tesauro		
	Documentos relevantes recuperados	Precisión	Exhaustividad	Documentos relevantes recuperados	Precisión	Exhaustividad
1.	26	1	0'63	25	1	0'60
2.	2	1	0'50	4	1	1
3.	5	0'55	1	3	0'60	0'60
4.	2	1	0'25	3	1	0'37
5.	19	0'34	0'65	13	1	0'44
6.	74	1	0'73	71	1	0'70
7.	4	1	0'30	4	1	0'30
8.	7	0'87	0'61	11	1	0'84
9.	1	1	0'09	9	1	0'81
10.	2	1	0'66	2	1	0'66

**Diagrama de barras representativo de la media de los índices en los distintos tipos de lenguajes:**



Como puede observarse, el lenguaje empleado en la formulación de las ecuaciones de búsqueda es lenguaje controlado técnico, dado que las entradas aceptadas por el tesoro se expresan en lengua latina. Paradójicamente, al recuperar en lenguaje libre aparecen los mismos documentos, por regla general, que con tesoro, ello se debe a que en el título de los documentos y en otros campos figuran los términos vulgares, así como por el hecho de que al truncar las palabras utilizadas en las ecuaciones usando el lenguaje natural, existe una coincidencia semántica parcial que favorece la concordancia de resultados.

En las ecuaciones en lenguaje natural incluimos todos los sinónimos y cuasisinónimos posibles para evitar omisiones, siendo más cómodo (por no necesitar tener en cuenta otros términos) el uso del lenguaje controlado. Tuvimos, por ejemplo que pensar en *células, huevos, embriones*, etc.. para sustituir al término preferente del tesoro *citodiferenciación*.

En la consulta número uno, formulada en lenguaje controlado, se perdió información debido a que el topónimo utilizado en la indización era más específico que el genérico *España* empleado en las ecuaciones de búsqueda. En las referencias encontramos, por ejemplo, *Río Guadiana, Costa Cantábrica, Estrecho de Gibraltar*, etc... Los documentos en los cuales se encontraban son obviamente relevantes, pero no se recuperaron con la estrategia por la razón mencionada.

En la segunda consulta, los 6 documentos recuperados con los dos tipos de lenguajes son relevantes, pero no son coincidentes. Los dos de libre son distintos a los cuatro de controlado, quizá porque la búsqueda en lenguaje técnico recupera más (4 sobre 2 de libre) pero no tiene en cuenta términos vulgares que aparecen en libre.

El nivel de precoordinación del tesoro, del 1'36%, tal y como veíamos en el capítulo anterior, permite precisar conceptos que, en libre, han de tratar de recuperarse mediante unitérminos o palabras clave, con un nivel de generalidad que, obviamente produce ruido o imprecisión, como en este caso de la quinta consulta, en el que la elección entre *ecología, bosques, y sistemas forestales* o el descriptor precoordinado *ecosistemas forestales*, hace variar el resultado en una proporción de 55/13 documentos recuperados. La sintagmatización de los descriptores del tesoro hace igualmente innecesarias un gran número de operaciones lógicas que, si bien restringen el volumen de items recuperados, provocan a veces pérdida de información pertinente.

Algunos documentos no se recuperaron porque estaban indizados por descriptores específicos, como, por ejemplo, los que respondían a la sexta pregunta: *rhinocerontidae*, *fósiles humanos* y *mammalia*, y en la ecuación de búsqueda se empleó el término genérico *fósiles vertebrados*.

Se dieron casos en los que entre las referencias aparecieron documentos que no estaban relacionados con el tema, al darse la coincidencia de que, tanto el descriptor, como la palabra clave se utilizaron en esta área de conocimiento con distintas connotaciones. El descriptor *dorada*, por ejemplo, aparece en asuntos relacionados con la vitivinicultura, en protección de la *Costa Dorada*, y en relación con la minería, en el artículo titulado: *La mina dorada: Un ejemplo de mineralización filoniana en un terreno metamórfico ligado a un magmatismo leucogranítico*. Estas imprecisiones se evitaron en lenguaje controlado al utilizar el término preferente *sparus aurata*, que resuelve la polisemia.

En ocasiones, y en contra de lo esperado, se consiguieron niveles de exhaustividad muy bajos, debido al escaso número de documentos que respondieron a los términos utilizados en las ecuaciones en lenguaje libre, por ejemplo, *anidamento* y *nidos*, en la consulta número nueve, recuperaron concretamente 36 items, mientras que con la entrada normalizada en el tesoro *nidificación*, se obtuvieron 114.

Se dió un único caso, concretamente en la consulta número diez, en el que coincidieron el número de registros y los documentos recuperados con ambos lenguajes.

### 6.6.2. Base de datos PIE (Programa de Investigación y Desarrollo Electrotécnico)

#### Consulta 1. *Sistemas de seguridad en centrales nucleares*

##### Búsqueda en lenguaje libre

S1 [sistema*]	630
S2 [seguridad]	128
S3 [central*]	430
S4 [nuclear*]	106
S5 [S1 Y S2]	73
S6 [S3 Y S4]	89
S7 [S5 Y S6]	12

$$\text{precisión} = \frac{7}{12} = 0'58$$

$$\text{exhaustividad} = \frac{7}{11} = 0'63$$

##### Búsqueda en lenguaje controlado

S1 [seguridad nuclear]	11
S2 [mantenimiento nuclear]	2
S3 [S1 O S2]	13

$$\text{precisión} = \frac{7}{13} = 0'53$$

$$\text{exhaustividad} = \frac{10}{11} = 0'90$$

## Consulta 2. Minicentrales de producción hidráulica

### Búsqueda en lenguaje libre

S1 [central o minicentral]	438
S2 [producción]	150
S3 [hidráulic*]	156
S4 [S2 Y S3]	42
S5 [S1 Y S4]	27

$$\text{precisión} = \frac{5}{27} = 0'18$$

$$\text{exhaustividad} = \frac{5}{28} = 0'17$$

### Búsqueda en lenguaje controlado

S1 [minicentral hidráulica]	28
-----------------------------	----

$$\text{precisión} = \frac{28}{28} = 1$$

$$\text{exhaustividad} = \frac{28}{28} = 1$$

### Consulta 3. Centrales de energía solar

#### Búsqueda en lenguaje libre

S1 [energía* Y sol*]	57
S2 [central]	422
S3 [S1 Y S2]	17

$$\text{precisión} = \frac{4}{17} = 0'23$$

$$\text{exhaustividad} = \frac{4}{9} = 0'44$$

#### Búsqueda en lenguaje controlado

S1 [energía solar]	33
S2 [central de energía]	20
S3 [S1 Y S2]	20

$$\text{precisión} = \frac{4}{20} = 0'20$$

$$\text{exhaustividad} = \frac{4}{9} = 0'44$$

### Consulta 4. Sensores ópticos

#### Búsqueda en lenguaje libre

S1 [sensor*]	41
S2 [óptico*]	13
S3 [S1 Y S2]	3

$$\text{precisión} = \frac{3}{3} = 1$$

$$\text{exhaustividad} = \frac{3}{3} = 1$$

**Búsqueda en lenguaje controlado**

**S1 [sensor óptico] 2**

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{3} = 0'66$$

**Consulta 5. Recursos de agua en forma de nieve**

**Búsqueda en lenguaje libre**

**S1 [agua o nieve] 185**  
**S2 [recurso\*] 69**  
**S3 [S1 Y S2] 17**

$$\text{precisión} = \frac{2}{17} = 0'11$$

$$\text{exhaustividad} = \frac{2}{4} = 0'50$$

**Búsqueda en lenguaje controlado**

<i>S1</i> [recurso de nieve]	2
<i>S2</i> [agua de nieve]	1
<i>S3</i> [ <i>S1</i> O <i>S2</i> ]	3

$$\text{precisión} = \frac{2}{3} = 0'66$$

$$\text{exhaustividad} = \frac{2}{4} = 0'50$$

**Consulta 6. Análisis del rendimiento de turbinas**

**Búsqueda en lenguaje libre**

<i>S1</i> [turbina*]	54
<i>S2</i> [análisis*]	356
<i>S3</i> [rendimiento]	100
<i>S4</i> [ <i>S2</i> Y <i>S3</i> ]	31
<i>S5</i> [ <i>S1</i> Y <i>S4</i> ]	9

$$\text{precisión} = \frac{3}{9} = 0'33$$

$$\text{exhaustividad} = \frac{3}{5} = 0'60$$

Búsqueda en lenguaje controlado

S1 [turbina]	11
S2 [rendimiento energético]	8
S3 [central hidráulica]	20
S4 [análisis]	35
S5 [ensayo]	50
S6 [S4 O S5]	80
S7 [S1 Y S2]	3
S8 [S6 Y S7]	1

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{3}{5} = 0'60$$

**Consulta 7. Construcción de laboratorios de ensayo**

Búsqueda en lenguaje libre

S1 [construcción*]	147
S2 [planificación*]	87
S3 [instalación*]	354
S4 [ensay*]	249
S5 [laboratorio*]	116
S6 [S1 O S2]	225
S7 [S3 O S5]	429
S8 [S4 Y S7]	131
S9 [S6 Y S8]	40

$$\text{precisión} = \frac{9}{40} = 0'22$$

$$\text{exhaustividad} = \frac{9}{11} = 0'81$$

**Búsqueda en lenguaje controlado**

S1 [construcción]	40
S2 [laboratorio de ensayo]	24
S3 [S1 Y S2]	11

$$\text{precisión} = \frac{8}{11} = 0'72$$

$$\text{exhaustividad} = \frac{8}{11} = 0'72$$

**Consulta 8. Sistemas expertos utilizados en líneas**

**Búsqueda en lenguaje libre**

S1 [sistema* Y experto*]	53
S2 [línea*]	165
S3 [S1 Y S2]	4

$$\text{precisión} = \frac{3}{4} = 0'75$$

$$\text{exhaustividad} = \frac{3}{5} = 0'60$$

Búsqueda en lenguaje controlado

S1 [sistema experto]	22
S2 [línea de alta tensión]	9
S3 [red eléctrica de alta tensión]	10
S4 [S2 O S3]	18
S5 [S1 Y S4]	2

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{5} = 0'40$$

**Consulta 9. Defectos o problemas generados en la red eléctrica**

Búsqueda en lenguaje libre

S1 [avería*]	35
S2 [interferen*]	5
S3 [problem*]	155
S4 [defecto*]	27
S5 [apagón*]	3
S6 [red*]	457
S7 [electr*]	607
S8 [S1 O S2 O S3 O S4 O S5]	211
S9 [S6 Y S7]	286
S10 [S8 Y S9]	64

$$\text{precisión} = \frac{19}{64} = 0'29$$

$$\text{exhaustividad} = \frac{19}{19} = 1$$

### Búsqueda en lenguaje controlado

S1 [perturbación eléctrica]	12
S2 [perturbación de la red]	9
S3 [defecto de línea eléctrica]	0
S4 [red eléctrica]	47
S5 [S1 O S2]	14
S6 [S4 Y S5]	6

$$\text{precisión} = \frac{6}{6} = 1$$

$$\text{exhaustividad} = \frac{6}{19} = 0'31$$

### **Consulta 10. Almacenamiento de energía por superconductividad**

#### Búsqueda en lenguaje libre

S1 [energía*]	393
S2 [almacena*]	71
S3 [superc* O conduct*]	43
S4 [S1 Y S2]	44
S5 [S3 Y S4]	5

$$\text{precisión} = \frac{4}{5} = 0'80$$

$$\text{exhaustividad} = \frac{4}{5} = 0'80$$

Búsqueda en lenguaje controlado

S1 [superconductividad]	6
S2 [almacenamiento]	5
S3 [S1 Y S2]	4

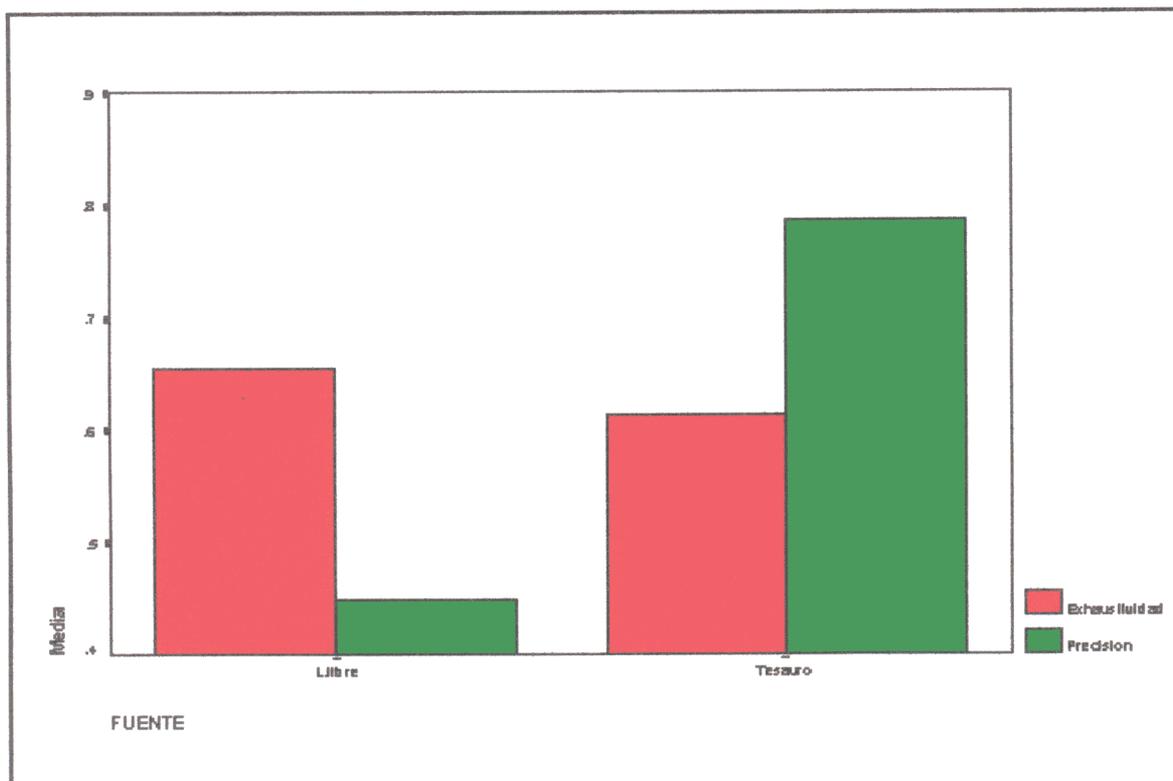
$$\text{precisión} = \frac{3}{4} = 0'75$$

$$\text{exhaustividad} = \frac{3}{5} = 0'60$$

**Cuadro comparativo de resultados:**

N° de Búsquedas	Lenguaje libre			Tesouro		
	Documentos relevantes recuperados	Precisión	Exhaustividad	Documentos relevantes recuperados	Precisión	Exhaustividad
1.	7	0'58	0'63	7	0'53	0'90
2.	5	0'18	0'17	28	1'00	1'00
3.	4	0'23	0'44	4	0'20	0'44
4.	3	1'00	1'00	2	1'00	0'66
5.	2	0'11	0'50	2	0'66	0'50
6.	3	0'33	0'60	1	1'00	0'60
7.	9	0'22	0'81	8	0'72	0'72
8.	3	0'75	0'60	2	1'00	0'40
9.	19	0'29	1'00	6	1'00	0'31
10.	4	0'80	0'80	3	0'75	0'60

Diagrama de barras representativo de la media de los índices en los distintos tipos de lenguajes:



En esta base de datos, la recuperación con ayuda del tesoro no fue, en ocasiones, lo efectiva que pudiera haberlo sido por el hecho -señalado en el capítulo quinto de esta investigación- de que no se incluyen descriptores de carácter general, capaces de representar conceptos amplios. En la segunda de las consultas de esta base, por ejemplo, fue preciso excluir de la estrategia el descriptor *producción de energía eléctrica*, presente en el tesoro, porque el resultado de su intersección con *minicentral eléctrica* era de 0 documentos. Probablemente el uso de un descriptor más genérico como el unitérmino

*producción* habría propiciado resultados óptimos en este caso; pero el tesoro excluye sistemáticamente la inclusión de términos generales, como decíamos anteriormente, incluyendo, sin embargo, entradas de extraordinario nivel de especificidad, lo cual beneficia la recuperación en casos concretos, pero la dificulta en otros, como el de la consulta de referencia.

Apareció un documento referido a *terreno*, que se recuperó al emplear el término polisémico *solar*. No obstante, se obtuvo en esta consulta prácticamente la misma tasa de precisión con ambos lenguajes, 0'23 y 0'20, e idéntica tasa de exhaustividad.

En lenguaje natural se obtuvo un documento más que, indizado bajo *sensor*, no se recuperó con el descriptor sintagmático *sensor óptico*. Podría deducirse de este caso que puede que el lenguaje libre dé mejores resultados cuando el proceso de indización se realiza a un nivel de especificidad bajo. El documento en sí trata de sensores en general y de sensores ópticos, aunque no lo hace de forma particular. En el supuesto de que hubieramos establecido baremos para estimar los índices de relevancia, este documento podría considerarse menos relevante que los otros dos recuperados en lenguaje controlado, pero como no se tuvo en cuenta esta variable, hemos de juzgarlo adecuado para la pregunta formulada.

El ruido producido en la quinta consulta en lenguaje libre, comparativamente al nivel de precisión conseguido en lenguaje controlado, se puede justificar por el desconocimiento del descriptor especializado *recurso de nieve*, que consiguió con mayor exactitud dar con las respuestas adecuadas.

En la séptima consulta se consiguió una baja precisión en lenguaje libre debido a la ambigüedad derivada de la combinación (intersección) de los unitérminos *ensayos*

y *laboratorio*, que dieron como resultado no sólo la obtención de documentos relativos a *laboratorios de ensayo*, sino los relacionados con *ensayos de laboratorio*. En lenguaje controlado, el descriptor sintagmático *laboratorio de ensayo* eliminó dicha ambigüedad.

Con la estrategia realizada mediante el tesauro se recuperaron todos los documentos relevantes de la colección, razón por la que coinciden los valores precisión y exhaustividad en este caso.

En la octava consulta se confirmó una vez más el hecho de que el uso de descriptores compuestos aporta mayor especificidad que el empleo de unitérminos.

Lo más significativo de la novena pregunta fue que el uso de sinónimos de la palabra *perturbación* (aceptada como entrada preferente en el tesauro) motivó un elevado volumen de referencias que, si bien perjudicaron la precisión, favorecieron la exhaustividad que, en este caso, resultó ser del 100%. Con tesauro, sin embargo, se logró un 100% de efectividad en el nivel de precisión y una baja tasa de exhaustividad, debido fundamentalmente al nivel de precoordinación de los descriptores y a la inexistencia de términos expresivos de conceptos tales como avería, interferencia, etc., presentes en los documentos recuperados mediante lenguaje libre.

### 6.6.3. Base de datos Biblioma, del Ministerio de Medio ambiente

**Consulta 1. Futuro de la producción de energía a partir de energías renovables.**

#### Búsqueda en lenguaje libre

1 [producción]	224
2 [generación]	257
3 [energía]	1308
4 [energías renovables]	215
5 [.L1 .O .L2 .Y .L3]	424
6 [.L5 .Y .L4]	35

$$\text{precisión} = \frac{30}{35} = 0'85$$

$$\text{exhaustividad} = \frac{30}{56} = 0'54$$

#### Búsqueda en lenguaje controlado

0 [generación de energía]	199
1 [energías renovables]	215
2 [.L0 .Y .L1]	31

$$\text{precisión} = \frac{30}{31} = 0'96$$

$$\text{exhaustividad} = \frac{30}{56} = 0'54$$

**Consulta 2. Contaminación de aguas subterráneas por residuos de plaguicidas.**

**Búsqueda en lenguaje libre**

0 [contaminación]	6008
1 [aguas subterráneas]	326
2 [pesticidas]	222
3 [insecticidas]	66
4 [plaguicidas]	29
5 [.L0 .Y .L1]	195
6 [(.L2 .O .L3 . OL4]	277
7 [.L6 .Y .L1]	12

$$\text{precisión} = \frac{11}{12} = 0'91$$

$$\text{exhaustividad} = \frac{11}{26} = 0'42$$

**Búsqueda en lenguaje controlado**

0 [contaminación de aguas subterráneas]	140
1 [pesticidas]	222
2 [.L0 .Y .L1]	11

$$\text{precisión} = \frac{11}{11} = 1$$

$$\text{exhaustividad} = \frac{11}{26} = 0'42$$

**Consulta 3. Influencia del turismo en el medio ambiente de las costas españolas.**

**Búsqueda en lenguaje libre**

0 [turis*]	490
1 [playas o costas]	5
2 [contamin*]	6323
3 [España]	2506
4 [.L1 .Y .L2 .Y .L3]	3

$$\text{precisión} = \frac{1}{3} = 0'33$$

$$\text{exhaustividad} = \frac{1}{2} = 0'50$$

**Búsqueda en lenguaje controlado**

0 [turismo]	484
1 [contaminación del litoral]	88
2 [España]	2506
3 [.L0 .Y .L1 .Y .L2]	1

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{2} = 0'50$$

#### Consulta 4. Impacto económico de la lluvia ácida

##### Búsqueda en lenguaje libre

0 [lluvia ácida]	205
1 [econom*]	4977
2 [.L0 .Y .L1]	14

$$\text{precisión} = \frac{7}{14} = 0'50$$

$$\text{exhaustividad} = \frac{7}{8} = 0'87$$

##### Búsqueda en lenguaje controlado

0 [lluvia ácida]	205
1 [economía]	3666
2 [.L0 .Y .L1]	5

$$\text{precisión} = \frac{5}{5} = 1$$

$$\text{exhaustividad} = \frac{5}{8} = 0'62$$

**Consulta 5. Incineración de residuos industriales tóxicos y/o peligrosos.**

**Búsqueda en lenguaje libre**

0 [incineración]	141
1 [residuos]	2228
2 [industri*]	2276
3 [toxic*]	707
4 [peligrosos]	332
5 [.L1 .Y .L2]	639
6 [.L3 .O .L4]	923
7 [.L0 .Y .L5 .Y .L6]	10

$$\text{precisión} = \frac{7}{10} = 0'70$$

$$\text{exhaustividad} = \frac{7}{39} = 0'17$$

**Búsqueda en lenguaje controlado**

0 [incineración]	141
1 [residuos peligrosos]	311
2 [residuos industriales]	336
3 [.L1 .Y .L2]	46
4 [.L0 .Y .L3]	4

$$\text{precisión} = \frac{4}{4} = 1$$

$$\text{exhaustividad} = \frac{4}{39} = 0'10$$

### **Consulta 6. Técnicas de reutilización y reciclado de neumáticos**

#### **Búsqueda en lenguaje libre**

0 [neumáticos]	12
1 [reutilización]	393
2 [.L0 .Y .L1]	11

$$\text{precisión} = \frac{11}{11} = 1$$

$$\text{exhaustividad} = \frac{11}{12} = 0'91$$

#### **Búsqueda en lenguaje controlado**

0 [neumáticos]	12
1 [reutilización de residuos]	366
2 [transformación de residuos]	107
3 [.L1 .O .L2]	452
4 [.L0 .Y .L3]	11

$$\text{precisión} = \frac{11}{11} = 1$$

$$\text{exhaustividad} = \frac{11}{12} = 0'91$$

**Consulta 7. Actividades agrarias lesivas para el medio natural**

**Búsqueda en lenguaje libre**

0 [agric*]	1513
1 [degradación]	522
2 [daños]	32
3 [deterioro]	24
4 [perjuicios]	9
5 [medio ambiente]	5803
6 [medio natural]	119
7 [.L1 .O .L2 .O .L3 .O .L4]	5888
8 [.L0 .Y .L7 .Y .L8]	23

$$\text{precisión} = \frac{18}{23} = 0'78$$

$$\text{exhaustividad} = \frac{18}{35} = 0'51$$

Búsqueda en lenguaje controlado

0 [agricultura]	541
1 [sobreexplotación del suelo]	5
2 [agotamiento de recursos]	28
3 [degradación de recursos naturales]	20
4 [destrucción de recursos naturales]	7
5 [degradación del suelo]	116
6 [deforestación]	79
7 [.L1 .O .L2 .O .L3 .O .L4 .O .L5 .O .L6]	244
8 [.L0 .Y .L7]	13

$$\text{precisión} = \frac{12}{13} = 0'92$$

$$\text{exhaustividad} = \frac{12}{35} = 0'34$$

**Consulta 8. Efectos de la radiactividad sobre el hombre**

Búsqueda en lenguaje libre

0 [radiactiv*]	417
1 [hombre]	495
2 [efectos]	2011
3 [.L0 .Y .L1 .Y .L2]	20

$$\text{precisión} = \frac{15}{20} = 0'7$$

$$\text{exhaustividad} = \frac{15}{15} = 1$$

Búsqueda en lenguaje controlado

0 [radiactividad o contaminación radiactiva]	165
1 [efectos sobre el hombre]	286
2 [.L0 .Y .L1]	15

$$\text{precisión} = \frac{15}{15} = 1$$

$$\text{exhaustividad} = \frac{15}{15} = 1$$

**Consulta 9. Polución por aceite**

Búsqueda en lenguaje libre

0 [polución]	16
1 [contaminación]	6008
2 [aceite]	12
3 [.L0 .O .L1 .Y .L2]	24

$$\text{precisión} = \frac{7}{24} = 0'29$$

$$\text{exhaustividad} = \frac{7}{7} = 1$$

Búsqueda en lenguaje controlado

0 [aceites y grasas]	15
1 [aceites usados]	25
2 [industria de aceites y grasas]	8
3 [contaminación]	4690
4 [.L0 .O .L1 .O .L2]	40
5 [.L3 .Y .L4]	7

$$\text{precisión} = \frac{6}{7} = 0'80$$

$$\text{exhaustividad} = \frac{6}{7} = 0'80$$

Consulta 10. Avances tecnológicos en la descontaminación atmosférica.

Búsqueda en lenguaje libre

0 [tecnologi*]	1784
1 [contaminación atmosférica]	1429
2 [control]	4600
3 [.L1 .Y .L2]	701
4 [.L0 .Y .L3]	77

$$\text{precisión} = \frac{26}{77} = 0'33$$

$$\text{exhaustividad} = \frac{26}{26} = 1$$

Búsqueda en lenguaje controlado

0 [control de la contaminación]	760
1 [desarrollo tecnológico]	81
2 [nuevas tecnologías]	48
3 [tecnología ambiental]	416
4 [contaminación atmosférica]	1163
5 [control ambiental]	518
6 [.L0 .O .L5]	1267
7 [.L1 .O .L2 .O .L3]	523
8 [.L4 .Y .L6 .Y .L7]	37

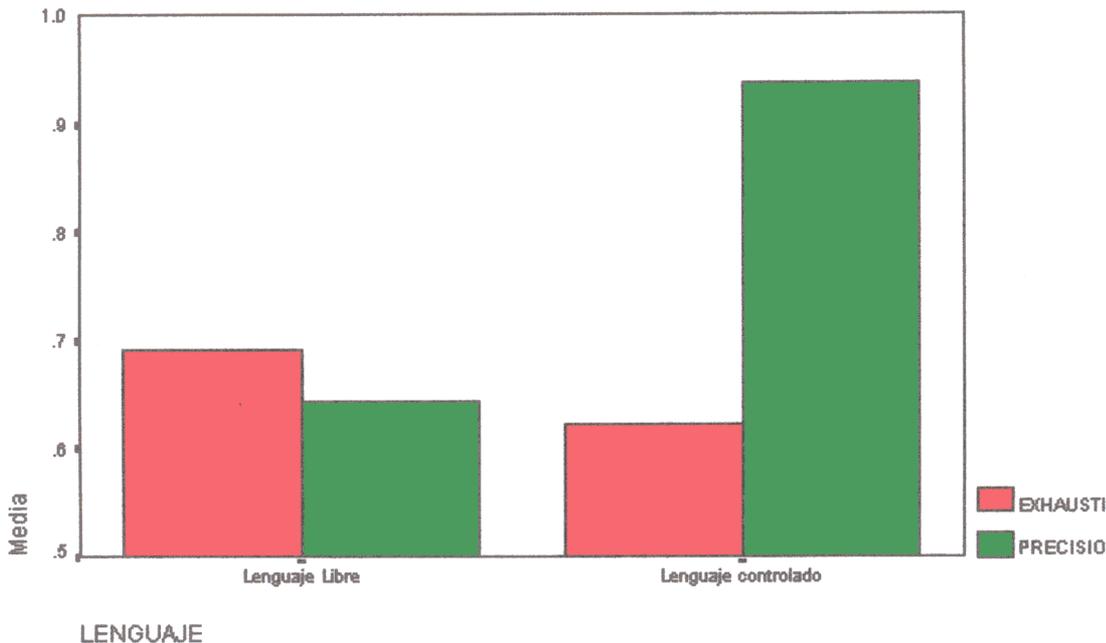
$$\text{precisión} = \frac{26}{37} = 0'70$$

$$\text{exhaustividad} = \frac{26}{26} = 1$$

**Cuadro comparativo de resultados:**

N° de Búsqueda	Lenguaje libre			Tesaurus		
	Documentos relevantes recuperados	Precisión	Exhaustividad	Documentos relevantes recuperados	Precisión	Exhaustividad
1.	30	0'85	0'54	30	0'96	0'54
2.	11	0'91	0'42	11	1	0'42
3.	1	0'33	0'50	1	1	0'50
4.	7	0'50	0'87	5	1	0'62
5.	7	0'70	0'17	4	1	0'10
6.	11	1	0'91	11	1	0'91
7.	18	0'78	0'51	12	0'92	0'34
8.	15	0'75	1	15	1	1
9.	7	0'29	1	6	0'80	0'80
10.	26	0'33	1	26	0'70	1

**Diagrama de barras representativo de la media de los índices en los distintos tipos de lenguajes:**



En las dos primeras consultas de esta base existe una coincidencia de términos empleados para formar las ecuaciones de búsqueda, y en consecuencia, los resultados en ambos procedimientos son similares.

En la cuarta ecuación, al truncar *econom\** aparecieron entre las referencias documentos no relevantes, entre otros uno cuyo resumen incluía la palabra Económica, referida a un organismo, pero que no respondía al perfil de la pregunta.

Los 4 documentos recuperados con lenguaje controlado en la quinta consulta son de alto nivel de relevancia. De los 10 recuperados en lenguaje libre, 9 son

relevantes, 4 de ellos coincidentes con los 4 conseguidos mediante tesauro, siendo útiles también los 5 restantes. La causa por la que parecen haberse perdido 5 referencias al formular la ecuación mediante tesauro puede deberse a que, en el momento de la indicación, no se caracterizaron los documentos por el descriptor *desechos industriales*, el cual, sin embargo, aparecía en los títulos de las revistas en las que se encontraban los artículos.

En búsquedas destinadas a usuarios potenciales, sería suficiente, en algunos casos, facilitar resultados obtenidos de forma bastante genérica, especialmente si el volumen de documentos no es muy amplio, así, por ejemplo, en la consulta número seis, la primera respuesta, con resultado de 12 documentos sobre el tema, probablemente sería suficiente para satisfacer sus necesidades, dada la escasez de referencias recuperadas, y el nivel de especialización de la base de datos en la que se realizó la operación. Sin embargo, dadas las características del proyecto que estamos desarrollando, preferimos confirmar la cantidad de registros que se obtienen en cada supuesto, procediendo, además, a la intersección con otros términos, que en el ejemplo de referencia correspondería a *reutilización*, descriptor que, como podemos comprobar, combinado con *neumáticos*, restringe solamente en uno el número de documentos relevantes hallados.

En casos aislados, en el listado de descriptores del CD-ROM no aparecieron descriptores que figuran en el tesauro impreso, por lo que fue necesario replantear la estrategia.

Como podremos observar en los comentarios realizados sobre las incidencias de las diferentes bases de datos consultadas, el recurso del truncamiento no siempre dió

los resultados apetecidos. En esta base en concreto, al truncar en sufijo *agri\** se recuperaron, entre otros, un documento no relevante en el que aparecía el Ministerio de Agricultura como editor, sin que dicho documento tenga relación alguna con el tema de la consulta.

En la octava estrategia podemos comprobar una vez más la importancia de la sintagmatización de los descriptores para obtener niveles de precisión altos, a expensas de la exhaustividad.

En la penúltima ecuación, el uso del término *aceite* en la estrategia en lenguaje libre originó considerable ruido documental, puesto que de los 24 documentos recuperados, sólo 7 resultaron relevantes, mientras que con los descriptores precoordiados del tesauro se consiguieron mejores niveles de precisión. No obstante, como es lógico, entre toda la información recuperada en lenguaje libre se encontraban todas las referencias relevantes de la base en lo que a *polución por aceite* se refiere.

De la última consulta podemos decir que el término *control de la contaminación* no está aceptado en el Tesauro de Medio Ambiente como descriptor. Sin embargo, al visualizar los registros, observamos que se utiliza en los procesos de indización como término preferente, por lo que tuvimos que hacer la unión de las entradas: *control de la contaminación* y *control ambiental*, no descriptor y descriptor respectivamente, para no perder información. Por lo que se refiere al lenguaje libre se dió el caso de que la intersección de *tecnologi\** y *contaminación atmosférica* produjo interferencias en los resultados, debido a las falsas combinaciones. Entre los documentos analizados se encontraban 26 relevantes que armonizaban con el tema de la consulta, y otros que trataban del impacto de las tecnologías en los procesos de contaminación atmosférica, que se oponía a la cuestión formulada.

#### 6.6.4. Base de datos del Centro de Documentación de la Mujer

**Consulta 1. Participación laboral de la mujer en el área de la ingeniería.**

**Búsqueda en lenguaje libre:**

\$1	participación laboral y ingeniería	1	(desc.)
\$2	*participación* y *ingen*	0	(tit.)
\$3	*participación* y *ingen*	6	(res.)

$$\text{precisión} = \frac{1}{6} = 0'16$$

$$\text{exhaustividad} = \frac{1}{10} = 0'10$$

**Búsqueda en lenguaje controlado:**

\$1	ingenieras	10
\$2	ingenieras y datos estadísticos	5

$$\text{precisión} = \frac{5}{5} = 1$$

$$\text{exhaustividad} = \frac{5}{10} = 0'50$$

**Consulta 2. Intervención de la mujer en el mundo sindical**

**Búsqueda en lenguaje libre**

§1	<b>*sindica*</b>	30 (en campo de descriptores)
§2	<b>*sindica* y mujer</b>	17 (en campo de título)
§3	<b>*sindica* y mujer</b>	194 (en campo de resu- men)

$$\text{precisión} = \frac{86}{194} = 0'44$$

$$\text{exhaustividad} = \frac{86}{93} = 0'92$$

**Búsqueda en lenguaje controlado**

§1	<b>sindicatos o sindicalistas o afiliación sindical</b>	82
----	---	----

$$\text{precisión} = \frac{56}{82} = 0'68$$

$$\text{exhaustividad} = \frac{56}{93} = 0'60$$

**Consulta 3. La alfabetización femenina en el medio rural**

Búsqueda en lenguaje libre

título:

\$1	*medio rural*	8
\$2	*mundo rural*	2
\$3	*pueblos*	3
\$4	*alfabetización femenina*	1
\$5	*alfabetización*	8
\$6	\$1 o \$2 o \$3	13
\$7	(\$4 o \$5) y mujeres	8
\$8	\$6 y \$7	21

resumen:

\$1	*medio rural*	37
\$2	*mundo rural*	20
\$3	*pueblos*	4
\$4	*alfabetización femenina*	1
\$5	*alfabetización*	34
\$6	\$1 o \$2 o \$3	85
\$7	(\$4 o \$5)	34
\$8	mujeres	5407
\$9	\$6 y \$7 y \$8	26

$$\text{precisión} = \frac{2}{26} = 0'07$$

$$\text{exhaustividad} = \frac{2}{8} = 0'25$$

Búsqueda en lenguaje controlado

\$1	alfabetización o analfabetas	34
\$2	mujer rural	163
\$3	acceso a la educación o educación de adultos	84
\$4	(\$1 o \$3) y \$2	5

$$\text{precisión} = \frac{4}{5} = 0'80$$

$$\text{exhaustividad} = \frac{4}{8} = 0'50$$

Consulta 4. Consumo de alcohol por adolescentes

Búsqueda en lenguaje libre

título:

\$1	*alcohol*	16
\$2	*adolescen*	52
\$3	*consumo de alcohol*	1
\$4	*bebidas alcohólicas*	1
\$5	\$1 o \$3 o \$4	16
\$6	\$5 y \$2	1

resumen:

\$1	*alcohol*	66
\$2	*adolescen*	134
\$3	*consumo de alcohol*	14
\$4	*bebidas alcohólicas*	3
\$5	\$1 o \$3 o \$4	66
\$6	\$5 y \$2	3

$$\text{precisión} = \frac{1}{4} = 0'25$$

$$\text{exhaustividad} = \frac{1}{5} = 0'20$$

Búsqueda en lenguaje controlado

\$1	alcohólicas	32
\$2	adolescentes	67
\$3	\$1 y \$2	2

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{5} = 0'40$$

**Consulta 5. Índices de participación de la mujer en el deporte en España.**

Búsqueda en lenguaje libre

título:

\$1	*deporte* o *deportista* o *gimnasia* o *gimnasta* o *aerobic*	38
\$2	\$1 o *ejercicio físico*	40

resumen:

\$1	*deport* o *ejercicio físico* o *gimnas* o *aerobic*	149
\$2	\$1 y *España*	26

$$\text{precisión} = \frac{29}{55} = 0'52$$

$$\text{exhaustividad} = \frac{29}{38} = 0'76$$

Búsqueda en lenguaje controlado

\$1	deporte	89
\$2	educación física	22
\$3	\$1 o \$2 y España	39

$$\text{precisión} = \frac{33}{39} = 0'84$$

$$\text{exhaustividad} = \frac{33}{38} = 0'86$$

Consulta 6. Sexismo en la escuela.

Búsqueda en lenguaje libre

título:

\$1	*sexismo* o *discriminación sexual*	32
\$2	\$1 y *escuela*	2

**resumen:**

\$1	*sexismo* o *discriminación sexual*	68
\$2	\$1 y *escuela*	27

$$\text{precisión} = \frac{27}{29} = 0'93$$

$$\text{exhaustividad} = \frac{27}{52} = 0'51$$

**Búsqueda en lenguaje controlado**

\$1	educación sexista	201
\$2	enseñanza preescolar o enseñanza de 1° grado o enseñanza de 2° grado o estudiantes de 1° grado o estudiantes de 2° grado	104
\$3	\$1 y \$2	33

$$\text{precisión} = \frac{33}{33} = 1$$

$$\text{exhaustividad} = \frac{33}{52} = 0'63$$

## Consulta 7. Imagen de la mujer en la publicidad

### Búsqueda en lenguaje libre

título:

\$1	*publici*	41
\$2	*televisi*	46
\$3	*medios de comunicación*	33
\$4	*anuncios*	2
\$5	*imagen de la mujer*	23
\$6	(\$1 o \$2 o \$3 o \$4) y \$5	7

resumen:

\$1	*publici*	142
\$2	*televisi*	135
\$3	*medios de comunicación*	249
\$4	*anuncios*	41
\$5	*imagen de la mujer*	101
\$6	(\$1 o \$2 o \$3 o \$4) y \$5	46

$$\text{precisión} = \frac{38}{45} = 0,84$$

$$\text{exhaustividad} = \frac{38}{73} = 0,52$$

### Búsqueda en lenguaje controlado

\$1	publicidad	90
\$2	imagen de la mujer	56

$$\text{precisión} = \frac{53}{56} = 0'94$$

$$\text{exhaustividad} = \frac{53}{73} = 0'72$$

**Consulta 8. El absentismo laboral de la mujer**

**Búsqueda en lenguaje libre**

título:

\$1	*absentismo laboral*	3
\$2	*baja por enfermedad*	0
\$3	*baja laboral*	0

resumen:

\$1	*absentismo laboral*	10
\$2	*baja por enfermedad*	0
\$3	*baja laboral*	1
\$4	\$1 o \$2 o \$3	11

$$\text{precisión} = \frac{9}{14} = 0'64$$

$$\text{exhaustividad} = \frac{9}{10} = 0'90$$

**Búsqueda en lenguaje controlado**

\$1	absentismo laboral	8
-----	--------------------	---

$$\text{precisión} = \frac{8}{8} = 1$$

$$\text{exhaustividad} = \frac{8}{10} = 0'80$$

**Consulta 9. Empleo de mujeres inmigrantes**

**Búsqueda en lenguaje libre**

título:

\$1	*inmigra*	22
\$2	*mujeres inmigrantes*	3
\$3	*dominicanas*	5
\$4	*empleo*	164
\$5	*trabajo*	406
\$6	*empleadas del hogar*	0
\$7	\$1 o \$2 o \$3	27
\$8	\$7 y (\$4 o \$5)	0

resumen:

\$1	*inmigra*	64
\$2	*mujeres inmigrantes*	17
\$3	*dominicanas*	12
\$4	*empleo*	909
\$5	*trabajo*	2888
\$6	*empleadas del hogar*	4
\$7	\$1 o \$2 o \$3	70
\$8	\$4 o \$5 o \$6	3304
\$9	\$7 y \$8	30

$$\text{precisión} = \frac{9}{30} = 0'30$$

$$\text{exhaustividad} = \frac{9}{12} = 0,75$$

Búsqueda en lenguaje controlado

\$1	migrantes	95
\$2	migración	35
\$3	acceso al empleo	79
\$4	empleo	29
\$5	servicio doméstico	42
\$6	\$1 o \$2	115
\$7	\$3 o \$4 o \$5	108
\$8	\$6 y \$7	3

$$\text{precisión} = \frac{0}{3} = 0$$

$$\text{exhaustividad} = \frac{0}{12} = 0$$

Consulta 10. **Feminismo en España**

Búsqueda en lenguaje libre

título:

\$1	*feminismo*	147
\$2	*movimiento feminista*	12
\$3	*liberación de la mujer*	9
\$4	(\$1 o \$2 o \$3) y España	54

resumen:

\$1	*feminismo*	547
\$2	*movimiento feminista*	201
\$3	*liberación de la mujer*	63
\$4	(\$1 o \$2 o \$3) y España	151

$$p = \frac{94}{151} = 0'62$$

$$\text{exhaustividad} = \frac{94}{148} = 0'63$$

Búsqueda en lenguaje controlado

\$1	feminismo	162
\$2	feminismo socialista	54
\$3	movimiento sufragista	76
\$4	movimiento feminista	239
\$5	feministas	115
\$6	\$1 o \$2 o \$3 o \$4 o \$5	536
\$7	\$6 y España	125

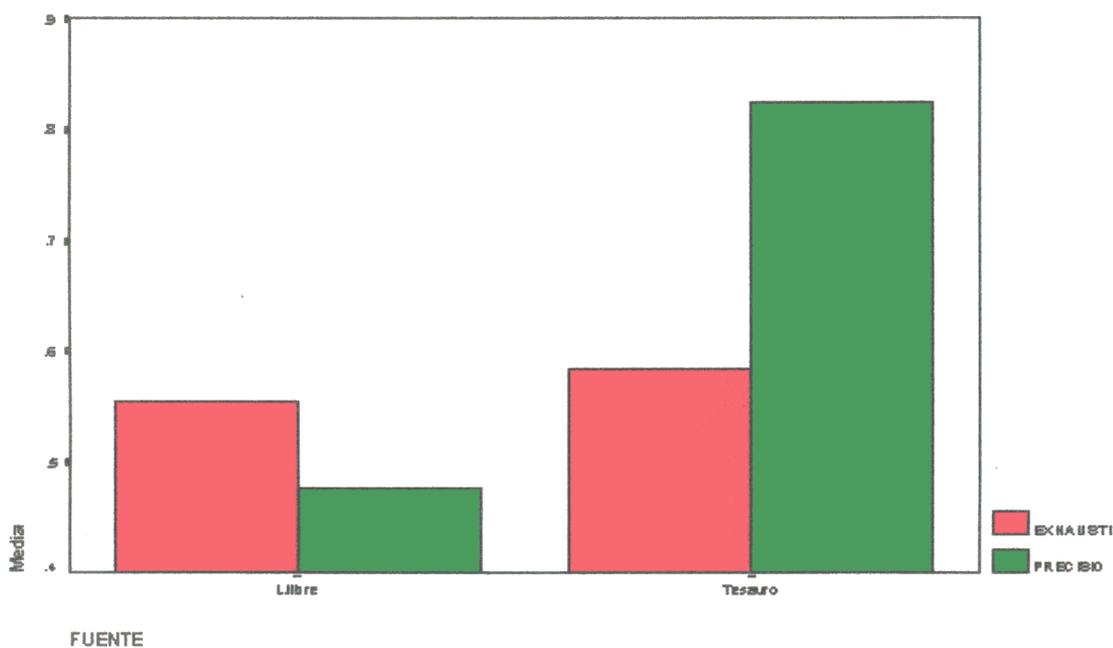
$$p = \frac{123}{125} = 0'98$$

$$\text{exhaustividad} = \frac{123}{148} = 0'83$$

**Cuadro comparativo de resultados:**

N° de Búsqueda	Lenguaje libre			Tesauro		
	Documentos relevantes recuperados	Precisión	Exhaustividad	Documentos relevantes recuperados	Precisión	Exhaustividad
1.	1	0'16	0'10	5	1	0'50
2.	86	0'44	0'92	56	0'68	0'60
3.	2	0'07	0'25	4	0'80	0'50
4.	1	0'25	0'20	2	1	0'40
5.	29	0'52	0'76	33	0'84	0'86
6.	27	0'93	0'51	33	1	0'63
7.	38	0'84	0'52	53	0'94	0'72
8.	9	0'64	0'90	8	1	0'80
9.	9	0'30	0'75	0	0	0
10.	94	0'62	0'63	123	0'98	0'83

**Diagrama de barras representativo de la media de los índices en los distintos tipos de lenguajes:**



Las estrategias en lenguaje libre en esta base de datos se realizaron en campos de título y resumen por separado por la imposibilidad de llevarlas a cabo de forma global, razón por lo que se incluyen ambas. Para hallar los índices de precisión y exhaustividad se analizaron y tuvieron en cuenta las referencias recuperadas en ambos campos, no siempre coincidentes, para comprobar el total de documentos resultantes de las búsquedas.

En la tercera estrategia en lenguaje libre, truncamos en prefijo y sufijo en el campo de título para que apareciera información, de lo contrario, sólo se recuperaba si coincidía el título con la palabra clave que se incluye en la ecuación, porque el ordenador lee secuencialmente. El resultado, en cualquier caso, mostró que en lenguaje natural, salvo dos documentos, los demás trataban el tema de forma marginal.

El truncamiento de sufijo *deport\**, en la quinta consulta, motivó la recuperación de un documento que trataba el tema de deportadas de guerra, que, naturalmente, no era relevante.

En la novena consulta, la incorporación a la estrategia del término *dominicanas*, en lenguaje libre, fue decisivo para hacer posible la recuperación de 9 documentos relevantes, que no se localizaron por medio del tesoro al no estar incluida esta palabra en su vocabulario. Este es probablemente el supuesto que muestra de forma más evidente cómo la facilidad de uso de términos nuevos en el vocabulario favorece la recuperación en lenguaje natural.

### 6.6.5. Base de datos PSEDISOC del CSIC

#### Consulta 1. *Desensibilización sistemática aplicada a fobias*

##### Búsqueda en lenguaje libre

1	42	'desensibilización'
2	56	'sensibilización'
3	9	'desensibilización' and ti 'sistemática'
4	14	'desensibilización' and re 'sistemática'
5	20	#3 or #4
6	98	#1 or #2 or #5
7	186	'fobias' or 'aversión' or 'temor'
8	9	#6 and #7

$$\text{precisión} = \frac{4}{9} = 0'44$$

$$\text{exhaustividad} = \frac{4}{5} = 0'80$$

##### Búsqueda en lenguaje controlado

1	212	'fobias' or de 'miedo' or de has 'aversión'
2	134	de '*fobia'
3	337	#1 or #2
4	31	de 'terapia de desensibilización sistemática'
5	5	#3 and #4

$$\text{precisión} = \frac{4}{5} = 0'80$$

$$\text{exhaustividad} = \frac{4}{5} = 0'80$$

**Consulta 2. Técnicas de relajación aplicadas a trastornos del sueño.**

**Búsqueda en lenguaje libre**

1	151	'relajacion'
2	428	'sueño' or 'descanso'
3	4024	'problemas*' or
4		'perturbaciones' or 'desordenes*'
5	2207	'trastornos*'
6	55	#13 and (#3 or #4)
7	1	#1 and #5
8	3	#1 and #2

$$\text{precisión} = \frac{2}{3} = 0'66$$

$$\text{exhaustividad} = \frac{2}{5} = 0'40$$

**Búsqueda en lenguaje controlado**

1	58	'tecnicas de relajacion'
2	30	'insomnio' or 'trastornos del sueño'
3	2	#1 and #2

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{5} = 0'40$$

**Consulta 3. El papel del psicólogo en los servicios sociales**

**Búsqueda en lenguaje libre**

1	7630	'papel' or 'rol' or 'funcion'
2	595	'psicologo?'
3	95	#1 and #2
4	1033	'servicios sociales'
5	7	#3 and #4

$$\text{precisión} = \frac{6}{7} = 0'85$$

$$\text{exhaustividad} = \frac{6}{62} = 0'09$$

**Búsqueda en lenguaje controlado**

1	532	'psicologo?' or 'rol del terapeuta'
2	954	'servicios sociales'
3	22	#1 and #2

$$\text{precisión} = \frac{16}{22} = 0'72$$

$$\text{exhaustividad} = \frac{16}{62} = 0'25$$

**Consulta 4. Peritaje de niños en casos de maltrato infantil**

**Búsqueda en lenguaje libre**

1	300	'peritaje' or 'peritacion' or 'dictamen'
2	30	'perito?'
3	321	#1 or #2
4	3741	'niños' or 'infancia'
5	1010	'maltrato', 'maltratamiento', 'daño', 'violencia', 'lesiones'
6	235	'victimas'
7	92	#4 and (#5 or #6)
8	74	'niños maltratados'
9	138	#7 or #8
10	2	#3 and #9

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{6} = 0'33$$

Búsqueda en lenguaje controlado

1	68	'niños', 'niños maltratados'
2	2151	'niños'
3	641	'violencia', 'violencia familiar'
4	2760	#2 or #3
5	32	#2 and #4
6	125	'niños maltratados' 'niños abandonados', 'abuso de niños'
7	14	'abuso sexual' and 'niños'
8	147	#5 or #6 or #7
9	1269	'medida', 'informe pericial'
5	7	#8 and #9

$$\text{precisión} = \frac{6}{7} = 0'85$$

$$\text{exhaustividad} = \frac{6}{6} = 1$$

**Consulta 5. Efectos de la fatiga en la conducción**

Búsqueda en lenguaje libre

1	67	'conduccion'
2	121	'fatiga', 'cansancio', 'agotamiento'
3	1	#1 and #2

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{2} = 0'50$$

Búsqueda en lenguaje controlado

1	216	'conductores', 'accidentes de trafico', 'seguridad vial'
2	381	'fatiga*', 'estres*'
3	2	#1 and #2

$$\text{precisión} = \frac{2}{2} = 1$$

$$\text{exhaustividad} = \frac{2}{2} = 1$$

**Consulta 6. Trastornos de la alimentación en adolescentes**

Búsqueda en lenguaje libre

1	891	'adolescentes'
2	2752	'trastornos*', 'desordenes', 'perturbaciones*', 'ansiedad'
3	873	'alimentacion'
4	11	#2 and #3
5	228	'bulimia', 'anorexia', 'inapetencia', 'apetito', 'hambre'
6	231	#4 or #5
7	22	#1 and #6

$$\text{precisión} = \frac{21}{21} = 1$$

$$\text{exhaustividad} = \frac{21}{43} = 0'48$$

Búsqueda en lenguaje controlado

1	721	'adolescentes', 'preadolescentes'
2	201	'nutricion', 'apetito', 'bulimia', 'anorexia nerviosa'
3	31	'trastornos de la conducta alimenticia'
4	3	#1 and #3
5	20	#1 and (#2 or #3)
6	2	'nutricion' and 'trastornos*'
7	32	#3 or #6
8	91	'nutricion'
9	110	#2 and not #8
10	20	#1 and (#3 or #9)

$$\text{precisión} = \frac{19}{19} = 1$$

$$\text{exhaustividad} = \frac{19}{43} = 0'44$$

**Consulta 7. Estrés o ansiedad en el contexto hospitalario**

**Búsqueda en lenguaje libre**

- |   |      |  |
|---|------|--|
| 1 | 1124 | 'estres', 'agotamiento',<br>'cansancio', 'ansiedad'        |
| 2 | 1116 | 'hospitales', 'clínicas',<br>'sanatorios', 'convalecencia' |
| 3 | 0    | 'quirófano', 'ambulancia'                                  |
| 4 | 47   | #1 and #2  |

$$\text{precisión} = \frac{16}{47} = 0'34$$

$$\text{exhaustividad} = \frac{16}{30} = 0'53$$

**Búsqueda en lenguaje controlado**

- |   |     |  |
|---|-----|--|
| 1 | 873 | 'hospitales', 'medio<br>hospitalario', 'pacientes<br>hospitalizados' |
| 2 | 114 | 'clínicas'   |
| 3 | 884 | 'estres', 'estres mental',<br>'ansiedad'                             |
| 4 | 23  | #3 and (#1 or #2)  |

$$\text{precisión} = \frac{16}{23} = 0'73$$

$$\text{exhaustividad} = \frac{17}{30} = 0'56$$

**Consulta 8. Estrategias meta-cognitivas en estudiantes universitarios.**

**Búsqueda en lenguaje libre**

1	504	'estudiantes universitarios', 'universitarios'
2	2748	'metacognicion', 'cognicion', 'conocimiento'
3	14	#1 and #2

$$\text{precisión} = \frac{6}{14} = 0'42$$

$$\text{exhaustividad} = \frac{6}{6} = 1$$

**Búsqueda en lenguaje controlado**

1	2222	'estudiantes', 'universidad'
2	1234	'metaconocimiento', 'conocimiento'
3	13	#1 and #2

$$\text{precisión} = \frac{4}{13} = 0'30$$

$$\text{exhaustividad} = \frac{4}{6} = 0'66$$

**Consulta 9. Evaluación de mecanismos de defensa con criterios del DSM.**

**Búsqueda en lenguaje libre**

1	136	'dsm', 'criterios del dsm'
2	212	'trastornos afectivos'
3	19	#1 and #2

$$\text{precisión} = \frac{11}{19} = 0'57$$

$$\text{exhaustividad} = \frac{11}{31} = 0'35$$

**Búsqueda en lenguaje controlado**

1	502	'trastornos afectivos', 'trauma emocional', 'depression mayor'
2	70	'dsm-III'
3	18	#1 and #2

$$\text{precisión} = \frac{14}{18} = 0'77$$

$$\text{exhaustividad} = \frac{14}{31} = 0'45$$

**Consulta 10. Factores psicosociales o trastornos de personalidad que influyen sobre el consumo en los jóvenes.**

Búsqueda en lenguaje libre

1	3982	'consumo', 'consumismo'
2	2294	'personalidad', 'rasgos de personalidad', 'características de la personalidad'
3	49	'factores psicosociales'
4	1327	'publicidad'
5	3658	#2 or #3 or #4
6	2030	'jóvenes', 'juventud'
7	11	#5 and #6 and #1

$$\text{precisión} = \frac{1}{11} = 0'09$$

$$\text{exhaustividad} = \frac{1}{1} = 1$$

Búsqueda en lenguaje controlado

1	508	'rasgos de personalidad', 'factores psicosociales', 'factores de personalidad'
2	1776	'publicidad', 'factores sociales'
3	2255	#1 or #2
4	101	'consumismo', 'comportamiento de los consumidores'
5	1654	'juventud', 'adolescencia'
6	1	#3 and #4 and #5

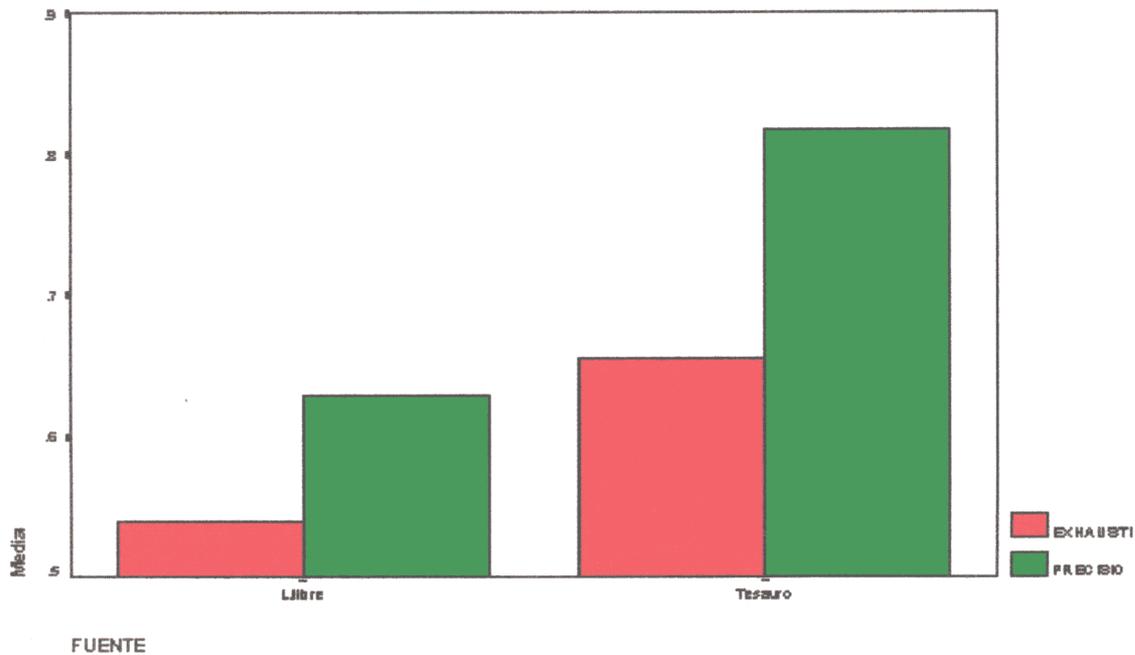
$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{1} = 1$$

**Cuadro comparativo de resultados:**

N° de Búsqueda	Lenguaje libre			Tesauro		
	Documentos relevantes recuperados	Precisión	Exhaustividad	Documentos relevantes recuperados	Precisión	Exhaustividad
1.	4	0'44	0'80	4	0'80	0'80
2.	2	0'66		2	1	
3.	6	0'85	0'09	16	0'72	0'25
4.	2	1	0'33	6	0'85	1
5.	1	1	0'50	2	1	1
6.	21	1	0'48	19	1	0'44
7.	16	0'34		0'17	0'73	
8.	6	0'42		4	0'30	
9.	11	0'57	0'35	14	0'77	0'45
10.	1	0'09	1	1	1	1

**Diagrama de barras representativo de la media de los índices en los distintos tipos de lenguajes:**



En líneas generales, la tasa de precoordinación del tesauro, 1'63, favoreció la precisión en lenguaje controlado. Por otra parte, la especificidad de algunas preguntas formuladas, propició la recuperación, mediante la estrategia original, de casi todos los documentos relevantes de la base sobre los temas de consulta. En las dos primeras preguntas, por ejemplo, el uso de descriptores precoordinados del tesauro restringió las respuestas, permitiendo un nivel de precisión mayor en lenguaje controlado que en libre, no observándose, además, mejores resultados en este último tipo de lenguaje en lo que a exhaustividad se refiere.

En respuesta a la tercera pregunta, la razón por la que se obtuvieron más referencias en lenguaje controlado, hecho poco habitual, fue la capacidad de inducción del tesaurus, que hizo posible la localización del descriptor *rol del terapeuta*. Este aportó la mayor parte de los documentos referidos al tema de consulta, con más exactitud que la palabra clave *psicólogo*, a pesar de estar ésta incluida en la pregunta.

En la base de datos había más documentos indizados bajo *metacognición* que bajo el descriptor aceptado en el tesaurus *metaconocimiento*, de ahí que, en la octava consulta, se recuperaran más referencias relevantes en lenguaje libre que por medio de la estrategia elaborada con el tesaurus.

En la última pregunta, la búsqueda por medio de los unitérminos *consumo* y *consumismo* dió como resultado 10 documentos irrelevantes, que no aparecieron cuando se concretó más la ecuación con el descriptor *comportamiento de los consumidores*, que recuperó un único documento totalmente relevante.

**6.6.6. Base de datos de la Biblioteca del Instituto de Migraciones y Servicios Sociales.**

**Consulta 1. Ayuda a familiares con enfermos de Alzheimer**

Búsqueda en lenguaje libre

\$1	69 respuesta(s) para:	AYUDA Y FAMILI*
\$2	178 respuesta(s) para:	ALZHEIMER
\$3	6 respuesta(s) para:	\$1 Y \$2

$$\text{precisión} = \frac{6}{6} = 1$$

$$\text{exhaustividad} = \frac{6}{21} = 0'28$$

Búsqueda en lenguaje controlado

\$1	172 respuesta(s) para:	ENFERMEDAD DE ALZHEIMER
\$2	20 respuesta(s) para:	SERVICIOS DE APOYO FAMILIAR
\$3	1 respuesta(s) para:	ENTRENAMIENTO EN HABILIDADES DE CUIDADO
\$4	3 respuesta(s) para:	ESTANCIAS TEMPORALES
\$5	6 respuesta(s) para:	GRUPOS DE APOYO
\$6	4 respuesta(s) para:	\$1 Y (\$2 o \$3 o \$4 o \$5)

$$\text{precisión} = \frac{4}{4} = 1$$



#TERCERA EDAD equivale a:

TERCERA EDAD O ANCIANOS ABANDONADOS O BONIFICACIONES A LA  
TERCERA EDAD O DERECHOS DEL ANCIANO O ENVEJECIMIENTO O  
ENVEJECIMIENTO BIOLÓGICO O ENVEJECIMIENTO PSICOLÓGICO O  
GERONTOLOGIA O JUBILADOS O PREPARACION A LA JUBILACION

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{2} = 0'50$$

Consulta 3. Trabajo en granjas escuela para niños deficientes psíquicos.

Búsqueda en lenguaje libre

\$1	102 respuesta(s) para:	CAMPO O GRANJA* O HORTICULT*
\$2	4981 respuesta(s) para:	DEFICIEN* O MINUSVAL* O DISCAPACI*
\$3	1208 respuesta(s) para:	TERAPIA
\$4	10 respuesta(s) para:	\$1 Y \$2
\$5	1 respuesta(s) para:	\$3 Y \$4

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{1} = 1$$

Búsqueda en lenguaje controlado

\$1	1429 respuesta(s) para:	DEFICIENCIA MENTAL
\$2	160 respuesta(s) para:	TERAPIA OCUPACIONAL
\$3	1 respuesta(s) para:	GRANJAS ESCUELA
\$4	21 respuesta(s) para:	\$1 Y \$2
\$5	0 respuesta(s) para:	\$3 Y \$4

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{1} = 1$$

**Consulta 4. Sexualidad en personas con deficiencia mental**

Búsqueda en lenguaje libre

\$1	509 respuesta(s) para:	SEX*
\$2	1774 respuesta(s) para:	(DEFIC* O MINUSVAL*) Y (MENTAL* O PSIQUIC*)
\$3	269 respuesta(s) para:	SUBNORMAL* O (RETRAS* Y MENTAL*)
\$4	1825 respuesta(s) para:	\$2 O \$3
\$5	72 respuesta(s) para:	\$1 Y \$4

$$\text{precisión} = \frac{44}{72} = 0'60$$

$$\text{exhaustividad} = \frac{44}{55} = 0'80$$

Búsqueda en lenguaje controlado

\$1	247 respuesta(s) para:	SEXUALIDAD
\$2	104 respuesta(s) para:	CONDUCTA SEXUAL
\$3	1596 respuesta(s) para:	#DEFICIENCIA MENTAL
\$4	305 respuesta(s) para:	\$1 O \$2
\$5	44 respuesta(s) para:	\$3 Y \$4

$$\text{precisión} = \frac{40}{44} = 0'90$$

$$\text{exhaustividad} = \frac{40}{55} = 0'72$$

**Consulta 5. Animales de compañía como medios terapéuticos en niños autistas.**

Búsqueda en lenguaje libre

\$1	337 respuesta(s) para:	AUTIS*
\$2	386 respuesta(s) para:	TERAPIA
\$3	6 respuesta(s) para:	ANIMALES O PERROS O GATOS
O		COBAYAS
\$4	5 respuesta(s) para:	\$1 Y \$2
\$5	1 respuesta(s) para:	\$3 Y \$4

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{3} = 0'33$$

\$1	5 respuesta(s) para:	ANIMALES DE COMPAÑIA
\$2	315 respuesta(s) para:	AUTISMO
\$3	79 respuesta(s) para:	NIÑOS CON NECESIDADES ESPECIALES
\$4	3 respuesta(s) para:	\$1 Y (\$2 O \$3)

$$\text{precisión} = \frac{3}{3} = 1$$

$$\text{exhaustividad} = \frac{3}{3} = 1$$

**Consulta 6. Avances genéticos en la prevención de deficiencias.**

**Búsqueda en lenguaje libre**

\$1	12 respuesta(s) para:	PREVENCIÓN Y GENÉTICA Y DEFICIEN*
-----	-----------------------	--------------------------------------

$$\text{precisión} = \frac{2}{12} = 0'16$$

$$\text{exhaustividad} = \frac{2}{3} = 0'66$$

**Búsqueda en lenguaje controlado**

\$1	58 respuesta(s) para:	PREVENCIÓN DE LA DEFI-
		CIENCIA
\$2	52 respuesta(s) para:	GENÉTICA
\$3	1 respuesta(s) para:	\$1 Y \$2

$$\text{precisión} = \frac{1}{1} = 1$$

$$\text{exhaustividad} = \frac{1}{3} = 0'33$$

**Consulta 7. Métodos de comunicación con discapacitados sensoriales.**

**Búsqueda en lenguaje libre**

\$1	141 respuesta(s) para:	SORDOS O CIEGOS O SORDO-
		MUDOS
\$2	1194 respuesta(s) para:	COMUNICACIÓN O LENGUAJE
\$3	49 respuesta(s) para:	\$1 Y \$2

$$\text{precisión} = \frac{26}{49} = 0'53$$

$$\text{exhaustividad} = \frac{26}{53} = 0'49$$

Búsqueda en lenguaje controlado

\$1	20 respuesta(s) para:	COMUNICACIÓN NO VOCAL
\$2	125 respuesta(s) para:	DEFICIENCIA SENSORIAL
\$3	282 respuesta(s) para:	SORDERA
\$4	241 respuesta(s) para:	DISCAPACIDAD
\$5	76 respuesta(s) para:	DISCAPACIDAD PARA LA COMUNICACIÓN
\$6	7 respuesta(s) para:	\$1 Y (\$2 O \$3 O \$4 O \$5)

$$\text{precisión} = \frac{7}{7} = 1$$

$$\text{exhaustividad} = \frac{7}{53} = 0'13$$

**Consulta 8. Actitudes sociales hacia los inmigrantes**

Búsqueda en lenguaje libre

\$1	158 respuesta(s) para:	INMIGRA*
\$2	79 respuesta(s) para:	EMIGRA*
\$3	71 respuesta(s) para:	EXTRANJEROS
\$4	32 respuesta(s) para:	COMPORTAMIENTO
\$5	738 respuesta(s) para:	CONDUCTA
\$6	254 respuesta(s) para:	ACTITUD
\$7	88 respuesta(s) para:	RACISMO
\$8	33 respuesta(s) para:	XENOFOBIA
\$9	225 respuesta(s) para:	MIGRA*
\$10	412 respuesta(s) para:	(\$1 O \$2 O \$3 O \$9)
\$11	1079 respuesta(s) para:	(\$4 O \$5 O \$6 O \$7 O \$8)
\$12	31 respuesta(s) para:	\$10 Y \$11

$$\text{precisión} = \frac{26}{31} = 0'83$$

$$\text{exhaustividad} = \frac{26}{32} = 0'81$$

Búsqueda en lenguaje controlado

\$1	111 respuesta(s) para:	ACTITUDES SOCIALES
\$2	32 respuesta(s) para:	XENOFOBIA
\$3	88 respuesta(s) para:	RACISMO
\$4	263 respuesta(s) para:	#MIGRACIÓN
\$5	20 respuesta(s) para:	TRABAJADORES EXTRANJEROS
\$6	5 respuesta(s) para:	TRABAJADORES MIGRANTES
\$7	231 respuesta(s) para:	(\$1 O \$2 O \$3)
\$8	274 respuesta(s) para:	(\$4 O \$5 O \$6)
\$9	22 respuesta(s) para:	\$7 Y \$8

$$\text{precisión} = \frac{16}{22} = 0'72$$

$$\text{exhaustividad} = \frac{16}{32} = 0'50$$

**Consulta 9. Problemas de alimentación en la tercera edad.**

Búsqueda en lenguaje libre

\$1	526 respuesta(s) para:	PROBLEMA*
\$2	48 respuesta(s) para:	DIFICULTAD*
\$3	947 respuesta(s) para:	TRASTORNO*
\$4	44 respuesta(s) para:	ALIMENT*
\$5	2563 respuesta(s) para:	TERCERA
\$6	2655 respuesta(s) para:	EDAD
\$7	806 respuesta(s) para:	ANCIAN*
\$8	30 respuesta(s) para:	VIEJ*
\$9	91 respuesta(s) para:	VEJ*
\$10	159 respuesta(s) para:	MAYORES
\$11	2549 respuesta(s) para:	\$5 Y \$6
\$12	1428 respuesta(s) para:	\$1 O \$2 O \$3
\$13	3143 respuesta(s) para:	\$7 O \$8 O \$9 O \$10
\$14	4 respuesta(s) para:	(\$4 Y \$12) Y \$13

$$\text{precisión} = \frac{1}{4} = 0'25$$

$$\text{exhaustividad} = \frac{1}{5} = 0'20$$

Búsqueda en lenguaje controlado

\$1	16 respuesta(s) para:	TRASTORNOS ALIMENTARIOS
\$2	14 respuesta(s) para:	DIETÉTICA
\$3	2905 respuesta(s) para:	TERCERA EDAD
\$4	573 respuesta(s) para:	ANCIANOS
\$5	3351 respuesta(s) para:	\$3 O \$4
\$6	30 respuesta(s) para:	\$1 O \$2
\$7	9 respuesta(s) para:	\$5 Y \$6

$$\text{precisión} = \frac{4}{9} = 0'44$$

$$\text{exhaustividad} = \frac{4}{5} = 0'80$$

**Consulta 10. El ejercicio físico en personas mayores.**

**Búsqueda en lenguaje libre**

\$1	198 respuesta(s) para:	DEPORT*
\$2	74 respuesta(s) para:	ACTIVIDAD
\$3	159 respuesta(s) para:	MAYORES
\$4	30 respuesta(s) para:	VIEJ*
\$5	91 respuesta(s) para:	VEJ*
\$6	2563 respuesta(s) para:	TERCERA
\$7	2655 respuesta(s) para:	EDAD
\$8	10 respuesta(s) para:	GIMNASIA
\$9	769 respuesta(s) para:	FISIC*
\$10	88 respuesta(s) para:	EJERCICI*
\$11	17 respuesta(s) para:	\$2 Y \$9
\$12	2549 respuesta(s) para:	\$6 Y \$7
\$13	31 respuesta(s) para:	\$9 Y \$10
\$14	219 respuesta(s) para:	\$1 O \$8 O \$11 O \$13
\$15	2585 respuesta(s) para:	\$3 O \$4 O \$5 O \$12
\$16	48 respuesta(s) para:	\$14 Y \$15

$$\text{precisión} = \frac{39}{48} = 0'81$$

$$\text{exhaustividad} = \frac{39}{52} = 0'75$$

Búsqueda en lenguaje controlado

\$1	2905 respuesta(s) para:	#TERCERA EDAD
\$2	573 respuesta(s) para:	ANCIANOS
\$3	169 respuesta(s) para:	ACTIVIDADES DEPORTIVAS
\$4	14 respuesta(s) para:	EJERCICIO FÍSICO
\$5	3351 respuesta(s) para:	\$1 O \$2
\$6	178 respuesta(s) para:	\$3 O \$4
\$7	49 respuesta(s) para:	\$5 Y \$6

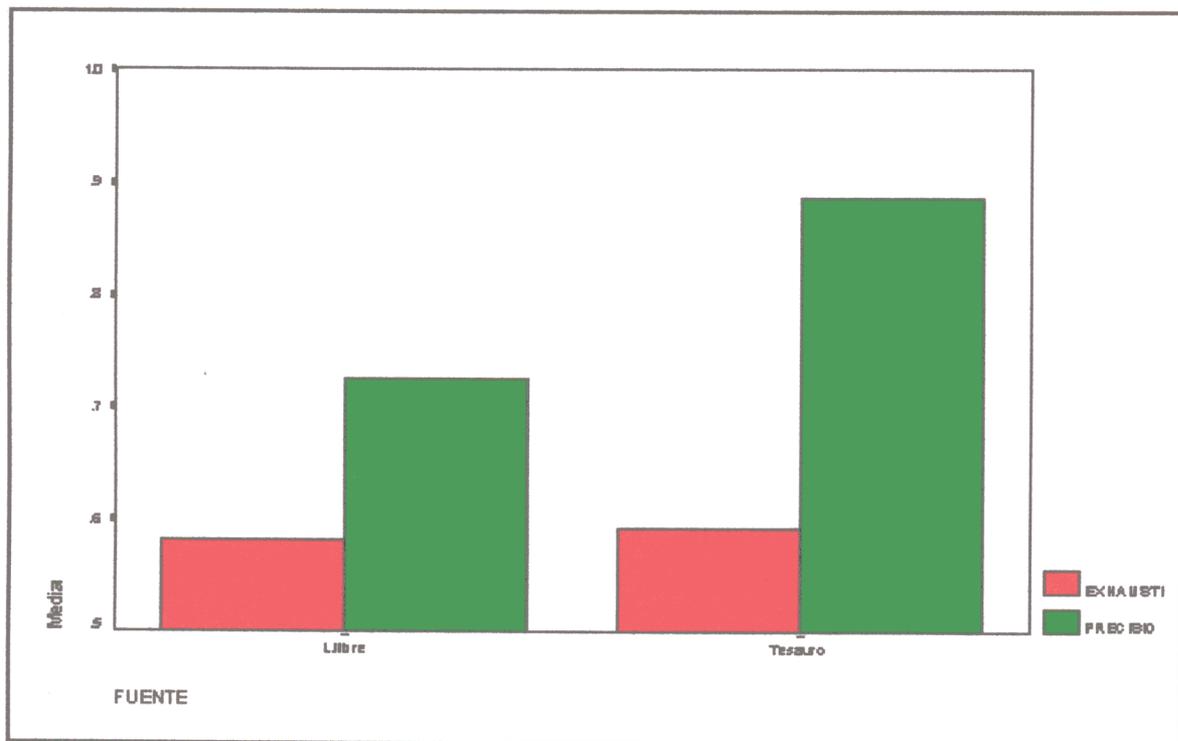
$$\text{precisión} = \frac{40}{49} = 0'81$$

$$\text{exhaustividad} = \frac{40}{52} = 0'76$$

**Cuadro comparativo de resultados:**

N° de Búsqueda	Lenguaje libre			Tesauro		
	Documentos relevantes recuperados	Precisión	Exhaustividad	Documentos relevantes recuperados	Precisión	Exhaustividad
1.	6	1	0'28	4	1	0'19
2.	1	1	0'5	1	1	0'5
3.	1	1	1	1	1	1
4.	44	0'61	0'80	40	0'90	0'72
5.	1	1	0'33	3	1	1
6.	2	0'20	0'66	1	1	0'33
7.	26	0'55	0'49	7	1	0'13
8.	26	0'83	0'81	16	0'72	0'50
9.	1	0'25	0'20	4	0'44	0'80
10.	39	0'81	0'75	40	0'81	076

**Diagrama de barras representativo de la media de los índices en los distintos tipos de lenguajes:**



Las cuestiones se formularon sobre temas específicos, lo que contribuyó a generar silencio en las consultas.

El uso de la almohadilla (recurso del signo #) permitió incluir todos los términos que dependían jerárquicamente del descriptor utilizado en la estrategia, por ejemplo, en los casos segundo y cuarto.

Al igual que en la base de datos de Biología animal, se dió un caso en este centro, concretamente el segundo, en el que existió coincidencia de resultados en los dos tipos de lenguajes.

En relación con la tercera consulta, podemos decir que en el tesauro no existe el descriptor *horticultura*, no hallándose ninguna referencia en el proceso de búsqueda con entradas más específicas del tesauro tales como *granjas escuela*. En lenguaje libre, en cambio, pudimos llegar hasta ese nivel de especificidad, recuperándose un documento en cuyo título figuraba el mencionado término.

En la misma consulta, en lenguaje controlado y como puede observarse en el desarrollo de la ecuación, al recuperar un documento bajo el descriptor *granjas escuela* no seguimos realizando operaciones de intersección (sets \$4 y \$5), porque perdíamos el único documento relevante recuperado. Fue posible evitar su omisión porque la pregunta era muy específica y, por tanto, fácilmente controlable.

En la octava pregunta, el descriptor compuesto *comunicación no vocal* restringió la cantidad de documentos recuperados, al contrario que *comunicación* o *lenguaje*, utilizados en la ecuación en lenguaje libre.

### **6.7. Resultados obtenidos**

Al proceder a la comparación de resultados se detectaron accidentes lingüísticos tales como polisemias, sinonimias, etcétera, que hemos comentado al finalizar el desarrollo de las estrategias de cada base de datos. Estas cuestiones lingüísticas explican, por sí solas las diferencias encontradas entre ambos procedimientos de búsqueda.

En total, las 60 búsquedas en lenguaje libre recuperaron 1.524 documentos, de los cuales 838 fueron relevantes. El mismo número de búsquedas utilizando lenguaje controlado dió como resultado un total de 983 referencias, 848 de las cuales fueron relevantes. Los conjuntos de documentos recuperados variaron en función del nivel de especificidad de la información requerida.

De las estrategias realizadas en lenguaje libre, poco más de la mitad dieron resultados relevantes, concretamente el 54'98%. En general la mayor parte de los documentos resultantes en lenguaje controlado respondieron al tópico de la consulta; en total el 86% se juzgaron relevantes.

Se puede decir que del análisis del test resulta lo siguiente: de cada 100 documentos recuperados en lenguaje libre, el 45% podrían ser eliminados por carecer de interés y que la proporción de documentos irrelevantes recuperados mediante tesauro es de tan sólo el 13'7%.

Estos datos confirman el alto porcentaje de ruido documental que genera el lenguaje natural en comparación con el controlado. La tasa de silencio, sin embargo, fue muy similar con ambos procedimientos de búsqueda; a lo largo del test, el total de documentos relevantes no recuperados con tesauro fue de 568, en libre la cifra ascendió a 573. La diferencia, de tan sólo 5 documentos, no es estadísticamente significativa. El silencio documental generado en la búsqueda con lenguaje controlado tuvo su origen en la ausencia de descriptores apropiados en casos muy concretos. Este fenómeno se dió, por ejemplo, en la consulta número 9 de la base de datos del Centro de Documentación de la Mujer, motivado por la falta de un término representativo de un concepto que ha cobrado significación en el entorno social en fecha posterior a la edición del tesauro. En la cuestión se solicitaba documentación acerca del *Empleo de mujeres inmigrantes*, tema

sobre el que no se halló ninguna referencia con la ayuda del tesauro, pero que, con la introducción de la palabra clave *dominicanas*, dió como resultado un total de 9 referencias.

Las tasas de precisión y exhaustividad conseguidas mediante lenguaje libre y con tesauro se pueden comparar en el siguiente cuadro:

	Lenguaje libre			Tesauro		
	Documentos relevantes recuperados	Precisión	Exhaustividad	Documentos relevantes recuperados	Precisión	Exhaustividad
Biología animal	142	87'6%	54'20%	145	96%	63'20%
Electrotecnia	59	44'90%	65'50%	63	78'60%	61'30%
Medio ambiente	133	64'40%	69'20%	121	93'80%	62'30%
Mujer	287	47'70%	55'40%	317	82'40%	58'40%
Psicología	70	63'70%	54'80%	85	81'70%	65'60%
Servicios sociales	147	72'50%	58'20%	117	88'70%	59'30%

Si hallamos la media de los seis sistemas de recuperación los resultados arrojan las siguientes cifras: en lenguaje libre se dió una precisión del 54'98% y una exhaustividad del 59'5%; con la ayuda del tesauro ambos índices mejoraron: la precisión alcanzó un 86'8% y la exhaustividad un 61'6%.

Las diferencias son del 23'4% en precisión, favorables a la recuperación por medio de tesauro, y del 2'1% en exhaustividad, también con la balanza a favor del

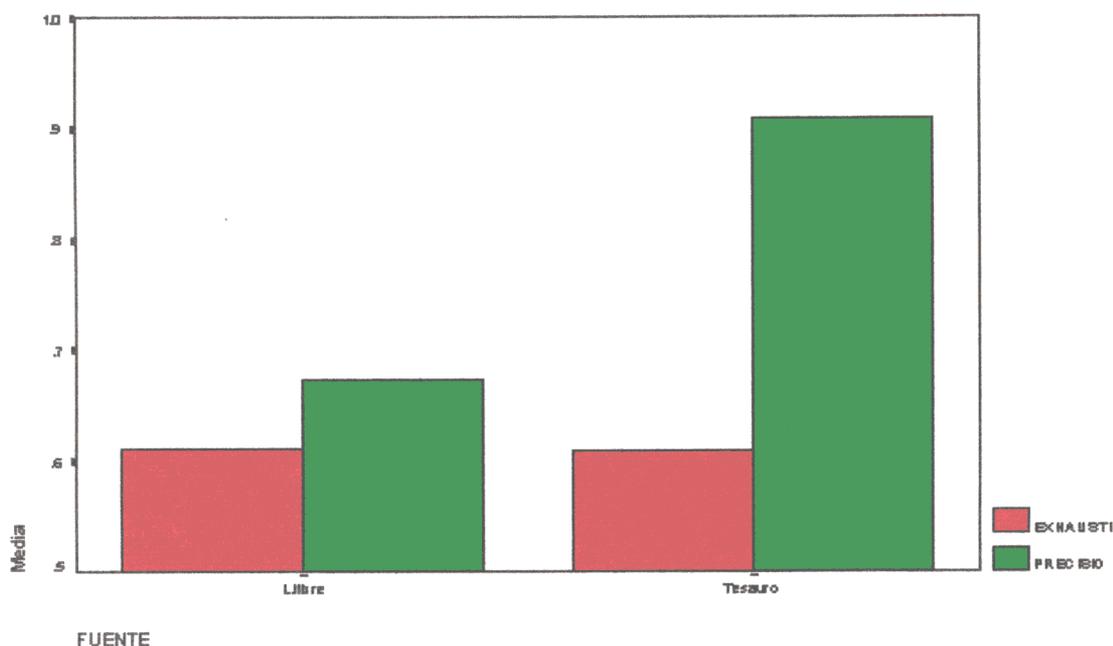
lenguaje controlado. Como puede observarse, uno de los resultados más significativos fue que los dos lenguajes proporcionan prácticamente el mismo nivel de rendimiento en lo que a exhaustividad se refiere, siendo, sin embargo, las diferencias registradas en las tasas de precisión estadísticamente significativas. Al comparar los valores de dichas tasas conseguidos en las tres bases de datos de ciencia y técnica con las de ciencias sociales no hallamos diferencias destacables. En lenguaje libre los valores de precisión en los dos tipos de bases alcanza una diferencia del 21'1%, y la exhaustividad un 20'3%, ambas favorables al campo de la técnica. La variación en los resultados obtenidos mediante tesoro es algo más significativa en lo que a índices medios de precisión se refiere (en los tres primeros servicios de recuperación se obtuvo un promedio del 23'83% y de un 22'96 en las tres últimas). En consecuencia, las diferencias muestran un 0'87% más de precisión en el conjunto de las bases de ciencias sociales, superando las del campo de la técnica a las sociales en tan sólo un 0'6% en lo que respecta a exhaustividad.

Los errores más comunes derivaron de las siguientes razones: a) inadecuada selección de términos de búsqueda; b) establecimiento de relaciones incorrectas entre los términos utilizados para preparar las ecuaciones; c) no tener en cuenta la existencia de polisemia en algunas materias para nosotros del todo desconocidas; d) incluir en las estrategias algún término incorrecto; e) representar algún concepto de forma demasiado amplia.

Como consecuencia, se dieron casos de irrelevancia, debido a que se utilizaron términos de búsqueda que representaban conceptos que no aparecían reflejados en las referencias recuperadas o que lo estaban de forma marginal. Los fallos fueron menos al emplear tesoros en las estrategias, gracias a que el nivel de precoordinación de todos ellos fue adecuado haciendo innecesario el uso de recursos sintácticos extra para unir los

términos precisos para la recuperación. Afortunadamente, fueron varios los casos en los que el cuidadoso repaso de las referencias permitió advertir fallos que no resultaban evidentes y que podrían haber modificado en alguna medida los resultados.

Mediante un diagrama de barras podemos observar de forma resumida y gráfica los valores conseguidos con ambos lenguajes:



Las columnas representan el promedio de cada índice en un tipo de lenguaje en particular, pudiendo observarse los siguientes aspectos:

Son patentes las diferencias existentes entre los valores recogidos de precisión y exhaustividad, siendo este último claramente inferior en ambos tipos de lenguajes. Concretamente, en libre la diferencia entre ambos valores es del 4'52% y en controlado del 25'2%. Aunque las cifras no lo reflejan, también es significativo que el

índice de exhaustividad sea superior con el uso del tesoro, lo cual contradice la creencia generalizada de que en lenguaje libre esta tasa es siempre mayor a la conseguida bajo control del vocabulario. Si bien a lo largo de la experimentación hubo casos en los que el uso del lenguaje libre dió un mayor índice de exhaustividad que el tesoro, fueron tan pocos que sus resultados no son estadísticamente significativos.

En resumen, el test tiende a mostrar que se obtienen mejores resultados en lenguaje controlado en lo que respecta a ambos índices.

## Conclusiones

## **Conclusiones**

En este trabajo hemos investigado acerca de los orígenes del empleo de los tesauros en España y establecido la comparación entre el uso del lenguaje libre y el controlado en la fase de recuperación documental. Los objetivos en este sentido fueron realizar una serie de observaciones o experimentos bajo condiciones reales en seis bases de datos, durante un período de diez meses. El análisis se centró en valorar las tasas de precisión y exhaustividad que permiten conseguir ambos tipos de lenguajes.

De los resultados de la investigación se pueden extraer las siguientes conclusiones:

1. Hasta 1982 no se empezaron a elaborar tesauros en España, existiendo hasta esa fecha un vacío prácticamente total en relación con su construcción y uso en sistemas documentales. Su desarrollo coincide con la creación de bases de datos, particularmente en el seno del CSIC, y la necesidad de conectar con Sistemas de Recuperación de Información internacionales.

2. El índice de utilización de tesauros en bases de datos españolas es bajo debido, a la falta de adecuación de dichos lenguajes documentales a necesidades específicas, o simplemente a su inexistencia en áreas temáticas especializadas. En los centros en los que existen, por ejemplo, listas de descriptores, desearían contar con un tesoro, pero no se construye porque requiere tiempo y esfuerzo económico que las bases no siempre están dispuestas a invertir.

3. Del análisis de los seis tesauros utilizados como muestra para comprobar su rendimiento se puede concluir que, cuanto más alto es su nivel de precoordinación, mayor es la diferencia en favor del tesoro (en relación con el lenguaje libre) en lo que a tasas de precisión se refiere. Igualmente, gracias a la tasa de precoordinación de los tesauros, que permitió definir los temas de las preguntas, se recuperaron menos documentos irrelevantes, es decir se generó menos ruido documental, que con el lenguaje libre.

4. La baja tasa de equivalencia de algunos tesauros -Electrotecnia (0'1) y Medio ambiente (0'26)- pudo haber influido en las tasas de exhaustividad, ya que son las dos únicas bases en las que el lenguaje libre supero al controlado en lo que a dicho valor se refiere.

5. El uso de un vocabulario inadecuadamente específico produce resultados con bajo nivel de relevancia y afecta a la exhaustividad en aquellos casos en los que no hay términos para describir conceptos significativos. De hecho, la ausencia de descriptores genéricos en algunos tesauros parece haber sido la causa por la cual el índice de exhaustividad utilizando estos lenguajes fue inferior al conseguido con lenguaje libre.

6. Los resultados de la comparación de la efectividad de seis tesauros y el lenguaje natural en el que se presentan los documentos de seis bases de datos bibliográficas demuestran que el control del vocabulario es factor determinante del éxito en el proceso de recuperación de información para lograr altos niveles de precisión; el aporte de sinónimos y otros términos relacionados con los conceptos de las preguntas mejoraron este índice en un 23'4% el conseguido mediante lenguaje natural.

7. El tesoro es el lenguaje de recuperación por excelencia en bases de datos especializadas, superando al libre incluso en lo que a índices de exhaustividad se refiere. Este hecho se da de forma particular en bases de ciencia y tecnología, caracterizadas por una terminología muy consolidada.

Si el lenguaje documental es completo, es decir, si incluye todos los posibles términos referidos a un mismo concepto, se consigue un elevado nivel de exhaustividad, en cuyo caso hay poca propensión por parte del usuario a utilizar en la recuperación términos no recogidos en su corpus. También puede influir en el mencionado resultado el nivel de consistencia en la indización (la utilización de los mismos términos siempre en el proceso), más probable en bases que emplean descriptores genéricos. La variable indización no ha sido, sin embargo, tenida en cuenta en este estudio, de modo que sólo la apuntamos como hipotética causa de influencia en los resultados.

Así pues, en contra de la creencia generalizada de que el empleo del lenguaje libre aporta más altos niveles de exhaustividad que el de los tesauros en la recuperación, con este lenguaje documental se consiguieron mejores tasas con el simple recurso de incorporar descriptores genéricos de partida y todos los específicos que comprenden. Paralelamente, en libre, la exhaustividad se incrementó usando todos los posibles sinónimos de los términos de búsqueda.

8. El lenguaje controlado consigue más bajos niveles de ruido que el libre, particularmente cuando los títulos de los documentos no son representativos de sus contenidos.

9. La tasa de silencio es muy similar con los procedimientos de búsqueda libre y controlada. El silencio documental generado en la búsqueda con tesaurus tuvo su origen en la ausencia de descriptores apropiados en casos muy concretos.

10. El control de la sinonimia en búsquedas en lenguaje libre obliga al buscador a realizar malabarismos lingüísticos para encontrar todas las posibles acepciones de un mismo concepto. Esta situación no se da con el uso de un tesaurus, en cuyo índice sistemático se encuentran todos los descriptores organizados por categoría semántica.

11. El uso del tesaurus permite contextualizar los términos de búsqueda evitando falsas combinaciones que surgen en lenguaje libre, particularmente cuando se han de emplear términos polisémicos en el proceso.

12. La preparación de ecuaciones se simplifica -cuando se utiliza un tesaurus- por su capacidad de incorporar a la búsqueda términos alternativos que concretan los temas propuestos en las cuestiones. Esa capacidad de inducción beneficia además la recuperación de documentos relevantes evitando un exceso de referencias inútiles cuyo examen lleva tiempo. Con lenguaje libre es el propio usuario quien ha de controlar el vocabulario cada vez que quiere recuperar información. De la experiencia aquí desarrollada se infiere que este hecho se da especialmente en bases de datos de ciencias sociales, en las que surgieron más casos de polisemia que en las bases técnicas.

13. El lenguaje libre ofrece la posibilidad de recuperar información muy especializada y actualizada, cuya terminología o no está incluida en el vocabulario controlado o no está representada de forma suficientemente específica.

14. Cuando las palabras del lenguaje natural coinciden con los descriptores o con los términos componentes de descriptores sintagmáticos, la recuperación con ambos procedimientos es equivalente, al igual que los resultados.

15. El análisis comparativo de uno y otro permite deducir que el lenguaje controlado neutraliza las deficiencias del lenguaje libre y viceversa, por lo que los dos sistemas no sólo no son antagonistas sino que se complementan uno al otro. La lista finita de descriptores de un tesoro no puede abarcar todos los conceptos de que se compone una disciplina y la inclusión de un campo adicional de palabras clave podría permitir la identificación de conceptos que no están recogidos en el corpus del tesoro. En muchas consultas (entre otros, este caso se dió, por ejemplo, en la base de datos ICYT de Biología, en la consulta referida a fauna marina) no coinciden los documentos recuperados por medio de ambos lenguajes, siendo, sin embargo, en ambos casos documentos relevantes los conseguidos, de ahí la conveniencia de utilizar los dos lenguajes en combinación para no perder información. Ello pone de manifiesto la inconsistencia en la indización en que, en mayor o menor grado, caen los analistas y productores de datos.

16. Los índices de precisión resultantes con ambos lenguajes tienen diferencias, en las cuales no ha influido el tipo de base de datos consultada, es decir, no se ha observado variación sustancial en lo que a este valor se refiere, si se comparan los resultados obtenidos en bases de datos de ciencia y tecnología o bases de ciencias sociales y humanidades.

### **Propuestas:**

- Programas automáticos internos de transformación del lenguaje libre en controlado (por el procedimiento de pregunta-respuesta). Como se procede en traducción

automática, que, por otra parte, debe estar incluida en los programas que proponemos, porque hay usuarios que buscan información en dos idiomas.

- Ofrecer al usuario formas de presentación de los lenguajes documentales más fáciles de usar, por ejemplo, interfaces de presentación de tesauros para usuarios más próximas al lenguaje natural, sin claves internas (TG, TE, etc.), o alternativas de presentación gráficas.

## Bibliografía

## **Capítulo 1. Objetivos, método, fuentes y bibliografía**

BOSCH GARCIA, C. **La técnica de investigación documental**. Caracas: Facultad de Ciencias Económicas, 1972.

CURRAS, E. **Documentación y metodología de la investigación científica. Cuaderno de trabajo**. Madrid: Paraninfo, 1985.

ECO, Umberto. **Cómo se hace una tesis. Técnicas y procedimientos de estudio, investigación y escritura**. 6ª ed. Barcelona: Gedisa, 1987.

LASSO DE LA VEGA JIMENEZ-PLACER, J. **Técnicas de investigación y Documentación. Normas y ejercicios**. Madrid: Paraninfo, 1980.

LOPEZ YEPES, J. **El estudio de la Documentación: bibliografía y metodología fundamentales**. Madrid: Tecnos, 1981.

----- **La aventura de la investigación científica: Guía del investigador y del director de investigación**. Madrid: Síntesis, 1995.

SIERRA BRAVO, R. **Tesis doctorales y trabajos de investigación científica**. Madrid: Paraninfo, 1994.

## **Capítulo 2. El tesoro en el marco de los lenguajes documentales.**

AITCHISON, J. Indexing languages. Classification schemes and thesauri. En L.J. Anthony, **Handbook of Special Libraries and information work**. 5ª ed. London: ASLIB, 1982, pp. 207-261.

AMAT NOGUERA, N. **Documentación científica y nuevas tecnologías de la información**. Madrid: Pirámide, 1987.

AMERICAN SOCIETY FOR TESTING AND MATERIALS (ASTM). **Thesaurus on resource recovery terminology**. Philadelphia: American Society for Testing and Materials, 1983.

AMERICAN NATIONAL STANDARDS INSTITUTE (ANSI). **Guidelines for thesaurus structure, construction and use, aproved July 10, 1973.** ANSI-239-19-1974. New York: ANSI, 1974.

AMERICAN NATIONAL STANDARDS INSTITUTE (ANSI). **Guidelines for thesaurus: structure, construction and use.** (Z 3919-1980). New York: ANSI, 1980.

ARTANDI, S. Coordinate indexing. En Kent, A. y Lancour, H. En: **Encyclopedia of Library and Information Science** . New York: Marcel Dekker, 1971, v. 5., pp. 679-682.

ASIS, M. Símbolos universales para su empleo en tesauros. **Revista Española de Documentación Científica**, v.2, 1979, pp. 39-47.

ASSOCIATION FRANÇAISE DE NORMALISATION (AFNOR). **Thesaurus monolingues et multilingues. Simbolisation des relations (norme experimental).** NF Z47-103. Paris: La Defense, 1978.

ASSOCIATION FRANÇAISE DE NORMALISATION (AFNOR). **Regles d'établissement des thesaurus en langue française.** Paris: AFNOR, 1973.

AUSTIN, D. Vocabulary control and information technology. **ASLIB Proceedings**, v. 38, n° 1 (1986) pp. 1-15.

AUSTIN, D. y WATERS, J. **Guidelines for the Establishment and Development of Multilingual Thesauri. Revised text.** London, Paris: UNESCO, The British Library, 1980.

AUSTIN, D. y DALE, P. **Guidelines for the establishment and development of monolingual thesauri.** 2<sup>a</sup> ed. rev. (PGI/81WS). Paris: UNESCO, 1981.

AUSTIN, D.; DALE, P.; UNESCO. **Directrices para el establecimiento y desarrollo de tesauros monolingües.** 2nd ed. rev. Paris: Unesco, 1984.

AVRAMESCU, A. Probabilistic criteria for the objective design of descriptor languages. **Journal of the American Society for Information Science.** v. 22, n° 1 (1971) pp. 85-95.

BLAGDEN, J. **Management Information Retrieval: A new indexing language.** London: British Institute of Management, 1969.

BRITISH STANDARD INSTITUTION (BSI). **Guidelines for the establishment and development of multilingual thesauri**. Norma 5723. London: BSI, 1979.

BRITISH STANDARD INSTITUTION (BSI). **Guidelines for the establishment and development of monolingual thesauri**. Norma 5723. London: BSI, 1979.

BUREAU MARCEL VAN DIJK. **Definition of Thesauri Essential Characteristics**. 2 vol. Brussels, 1976.

COYAUD, M. **Introduction à l'étude des langages documentaires**. Paris: Klincksieck, 1966.

CURRÁS, E. **Thesaurus. Lenguajes terminológicos**. Madrid: Paraninfo, 1991.

CHAUMIER, J. **Análisis y lenguajes documentales: El tratamiento lingüístico de la información documental**. Barcelona: Mitre, 1986.

CHAUMIER, J. y FOURTEAU, P. Le traitement des thesaurus a schemas fleches para l'informatique graphique interactive. **Documentaliste**, v. 16, n° 1 (1979) pp. 9-14.

CHEN, H. [et al]. Automatic thesaurus generation for an electronic community system. **Journal of the American Society for Information Science**, 46, 3, Apr (1995), pp. 175-93.

DAHLBERG, I. Concerning thesaurus terminology. **International Classification**, v. 2, nov. (1975) pp. 111-112.

----- The Broad System of Ordering (BSO) as a Basis for an Integrated Social Sciences Thesaurus?. **International Classification**, v. 7, n° 2 (1980) pp. 66-72.

DEVADASON, F.J. [et al]. Computer generation of thesaurus from structures subject-prepositions. **Information Processing & Management**, n° 17, (1981) pp. 1-11.

DIEMER, A. Classification, thesaurus and what next? The problem of the "third generation" in documentation and information. **Nachrichten für dokumentation**, v. 23 (1972) pp. 52-57.

DILLON, M.; CAPLAN, P.A. A technique for evaluating automatic terms clustering. **Journal of th American Society for Information Science**, 31, 2 (1980) pp. 89-96.

EUROPEAN ATOMIC ENERGY COMMUNITY (EURATOM). **EURATOM thesaurus.**, 2 v., 2ª ed. Bruselas: European Atomic Energy Community (EURATOM), 1967.

FOSKETT, D.J. Thesaurus. En **Encyclopedia of Library and Information Science**, v. 30. New York: Marcel Dekker, 1980, pp. 416-463.

FOSKETT, A.C. Postcoordinate indexing systems. En Foskett, A.C. **The subject approach to information**. London: Clive Bingley, 1996, pp. 379-440.

FREEDMAN, B. Thesauri for vocabulary control. **Drexel Library Quarterly**, v. 8 (1972) pp. 125-128.

FUGMANN, R. The glamour and the misery of the thesaurus approach, v. 4: On information retrieval theory. **International Classification**, v. 1, nº 2 (1974) pp. 76-86.

GANZMANN, J. Criteria for the evaluation of thesaurus software. **International Classification**, v. 17, nº 3-4 (1990) pp. 148-157.

GARCÍA GUTIÉRREZ, A. **Lingüística documental: Aplicación a la documentación de la comunicación social**. Barcelona: Mitre, 1984.

GARCIA GUTIERREZ, A. y LUCAS FERNANDEZ, R. **Documentación automatizada en los medios informativos**. Madrid: Paraninfo, 1987.

GILCHRIST, A. **The thesaurus in retrieval**. London: ASLIB, 1971.

----- The role of thesauri in mechanized systems. **Indexer**, nº 9, Oct. (1975) pp. 146-154.

HAYNES, D. Thesaurus construction using a micro. **Library Micromation News**, nº 13 (1986) pp. 8-12.

HINES, T. C. y HARRIS, J.L. Columbia University School of Library Service system for thesaurus development and maintenance. **Information Storage and Retrieval**, v. 7, nº 1 (1971) pp. 39-50.

INFOTERM. International Bibliography of Standardized and Non-Standardized Guidelines for Terminology. **International Classification**, v. 8, nº 2 (1981) pp. 98-100.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. (ISO). **Guidelines for the Establishment and Development of Monolingual Thesauri (2788)**. Geneva: ISO, 1974.

----- Directrices para el establecimiento y desarrollo de tesauros monolingües PNE 50.106, ISO 2788-1986 (E), parte 1, **Revista Española de Documentación Científica**, v. 12, nº 4 (1989) pp. 463-483.

----- Directrices para el establecimiento y desarrollo de tesauros monolingües PNE 50.106, ISO 2788-1986 (E), parte 2, **Revista Española de Documentación Científica**, v. 13, nº 1 (1989) pp. 601-629.

----- **Guidelines for the Establishment and Development of Multilingual Thesauri (5964)**. Geneva: ISO, 1985.

----- **Thesaurus ISO**. 3 v. Ginebra: ISO, AFNOR, 1976.

----- **ISONET indexing handbook**. Ginebra: ISO, 1978.

JEREZ, R. y RIOS, A. Los lenguajes documentales en España. **Documentación de las Ciencias de la Información**, nº 2 (1978) pp. 33-37.

JONES, K.S. Automatic thesaurus construction and the relation of a thesaurus to indexing terms. **ASLIB Proceedings**, v. 22, nº 5 (1970) pp. 226-228.

JONES, S. A thesaurus data model for an intelligent retrieval system. **Journal of Information Science**. v. 19, nº 3 (1993) pp. 167-178.

KAZLAUSKAS, E.J.; HOLT, T.D. The application of a minicomputer to thesaurus construction. **Journal of the American Society for Information Science**, 31, 5 (1980) pp. 363-368.

KLEINBART, P. Prolegomenon to "intelligent" thesaurus software. **Journal of Information Science**, 11, 2 (1985) pp. 45-53.

KROOKS, D.A. & LANCASTER, F.W. The evolution of guidelines for thesaurus construction. **Libri**, 43, 4, Oct-Dec (1993), pp. 326-42.

LANCASTER, F. W. **Vocabulary control for information retrieval**. Washington: Information Resource Press, 1972.

----- **Thesaurus construction and use: a condensed course.** Paris: Unesco, 1985.

----- **El control del vocabulario en la recuperación de información.** Valencia: Universitat, 1995.

LAUREILHE, M.T. **Le thesaurus, son rôle, sa structure, son élaboration.** 2<sup>a</sup> ed. Villeurbanne: Presses de l'E.N.S.B, 1981.

LORCH, W.T. International guidelines for thesauri: a description of the third Unesco draft. **Nachrichten für dokumentation**, v. 21 (1970) pp. 72-75.

MACCAFFERTY, M.A. **Thesauri and thesauri construction.** London: ASLIB, 1977.

MAIXNER, V. Towards compensation laws in constructing thesauri. **Information Storage and Retrieval**, v. 6, n° 4 (1970) pp. 383-385.

MANIEZ, J. **Lenguajes documentales y de clasificación: Concepción, construcción, y utilización en los sistemas documentales.** Madrid: Fundación Germán Sánchez Ruipérez, 1992.

NEELAMEGHAN, A. Non hierarchical associative relationships: their types and the computer generation of TR links. En: **Seminar on thesaurus in information systems.** Bangalore, December 1-5, 1975, pp. A1-A8.

RICHMOND, P.A. A thesaurus within a thesaurus: a study in ambiguity. En: **Towards a theory of librarianship: papers in honour of Jesse Hauk Shera.** (Conrad H. Rawski, ed.). Metuchen, N.J.: Scarecrow Press, 1973, pp. 268-301.

RODRIGUEZ DELGADO, R. **Uso de descriptores y tesauros.** Santiago de Chile: CEPAL-CLADES, 1971.

----- Teoría de sistemas y tesauros facetados. **Revista Española de Documentación Científica**, v. 1, n° 3 (1978) pp. 213-224.

----- Unificación de lenguajes documentarios. Bases para la creación de tesauros comunes en español. **Boletín de la ANABAD**, v. 28, n° 4 (1978) pp. 57-63.

----- Towards the integration of Decimal Classification and Thesauri concepts. En Taylor, P. J., editor. En: **New trends in documentation and information. Proceedings of the 39th FID Congress.** London: ASLIB, 1980, pp. 277-294.

----- La integración de los lenguajes documentarios, fin de Babel. **Revista Española de Documentación Científica**, v. 4, n° 3 (1980) pp. 333-341.

ROHOU, C. La gestion automatisée des thesaurus: étude comparative de logiciels. **Documentaliste**, v. 24, n° 3, May June (1987), pp. 103-108.

ROLLING, L.N. The role of graphic display of concept relationships in indexing and retrieval vocabularies. En: Classification research: **Proceedings of the second international study conference** (Pauline Atherton, ed.). Munksgaard, Copenhagen, 1965, pp. 295-320.

ROLLING, L.N. Compilation of thesauri for use in computer systems. **Information storage and retrieval**, n° 6 (1970) pp. 101-350.

SALTON, G. **Experiments in automatic thesaurus construction for information retrieval**. **Information storage and retrieval**. Ithaca, NY: Cornell University, Department of Computer Science, 1971.

SHEPHERD, M. & WATTERS, C. Computer generation of thesaurus. **Library Science with a Slant to Documentation** (Thesaurus Ser. 1), n° 12, June (1975) pp. 40-54.

SHILL, D.A. The role of a thesaurus in information retrieval. **Management Datamatics**, v. 4, Feb. (1975) pp. 9-15.

SLYPE, G. van. **Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales**. Madrid: Fundación Germán Sánchez Ruipérez, 1991.

SOERGEL, D. **Indexing languages and thesauri: Construction and maintenance**. Los Angeles: Melville, 1974.

SOMERS, H.L. Observations on standards and guidelines concerning thesaurus construction. **International Classification**, n° 2 (1981) pp.69-74.

TAUBE, M. Notes on the use of roles and links in coordinate indexing. **American Documentation**, v. 12, n° 2 (1961) pp. 98-100.

**TDCK Circular Thesaurus System**. The Hague: Wetenschappelijk en Technich Documentatie-en Informatiecentrum voor de Krijgsmacht, 1963.

**THESAURUS URBAMET.** 6ª ed. Paris: Ministère de l'Équipement des Transports et du Tourisme; Council Régional d'Île-de-France, 1993.

UNITED NATIONS ECONOMIC AND SOCIAL COUNCIL. **Use of descriptors and thesauri.** New York: United Nations Economic Commission for Latin America, 1971.

UNESCO; LANCASTER, L.W. **Thesaurus construction and use: A condensed course.** Paris: Unesco, 1985.

**URBAN Thesaurus.** Kent, Ohio: Kent State University, Center for Urban Regionalism, 1968.

**UNISIST Guidelines for the Establishment and development of Multilingual Thesauri (PGI-80/ws/12).** Revised edition. Paris: Unesco, 1980.

**UNISIST Guidelines for the Establishment and development of Monolingual Thesauri (PGI-81/ws/15).** Revised edition. Paris: Unesco, 1980.

WILLETTS, M. Investigation of the nature of the relation between terms in thesauri. **Journal of Documentation**, v. 31, n° 3, September (1975) pp. 158-184.

WOLFF-TERROINE, M. "A macrothesaurus. Why? How?". En: **Ordering Systems for Global Information Networks.** Edited by A. Neelameghan. Bangalore, India: International Federation for Documentation, 1979, pp. 431-436.

### **Capítulo 3. Evolución histórica de los tesauros**

AA, H.J. van der. **Thesaurus bibliography.** Amsterdam: Studiecentrum NOVI, 1977.

AARON, K. **Thesauri bibliography.** Haifa: Elyachar Library, Technion Israel Institute of Technology, 1975.

AENOR. **Documentación.** Tomo 2 Normas fundamentales. Madrid: Aenor, 1994.

AITCHISON, J. The EUDISED thesaurus: a preliminary appraisal. **Educational Libraries Bulletin**, v. 18 (1975) pp. 12-25.

----- **Tesauro de la Unesco**. 2 v. Paris: UNESCO, 1984.

----- Integration of thesauri in the Social Sciences. **International Classification**, v. 8, n° 2, pp. 75-85.

----- The Thesurofacet: a multipurpose retrieval language tool. **Journal of Documentation**, v. 26, n° 3 (1970) pp. 187-203.

AITCHISON, J. [et al]. **Thesurofacet: A thesaurus and faceted classification for engineering and related subjects**. Whetstone: English Electric Co., 1969.

AITCHISON, J. y ALLEN, C.G. **Bibliografía de vocabularios, tesauros, encabezamientos de materia y esquemas de clasificación de las Ciencias Sociales**. Paris: UNESCO, 1983.

AITCHISON, J, GOMERSALL, A. y IRELAND, R. **Thesurofacet: A thesaurus and faceted classification for engineering and related subjects**. Whetstone: The English Electric Company, 1969.

AITCHISON, J.; SMITH, I. A. y THOMPSON, S. **Thesaurus on Youth: an integrated classification and thesaurus for youth affairs and related topics**. Leicester: National Youth Bureau, 1981.

ALVES, S.C.; BASILIO, M.B.; BASTOS, S. **Tesauros, bibliografía, 1970-1982**. Brasilia: IBICT, 1983.

AMERICAN INSTITUTE OF CHEMICAL ENGINEERS. **Chemical engineering thesaurus; a wordbook for use with the concept coordination system of information storage and retrieval**. New York: American Institute of Chemical Engineers, 1961.

ARMED SERVICE TECHNICAL INFORMATION AGENCY, ARLINGTON, VIRGINIA. **Thesaurus of Astia Descriptors**. 2nd. ed. Arlington, Virginia: Astia, 1962.

ARNTZ, R. & PICHT, H. **Introducción a la terminología**. Madrid: Fundación Germán Sánchez Ruipérez, 1995.

ASSOCIATION FRANÇAISE DE NORMALISATION (AFNOR). **Thesaurus international technique. Noriane, Isonet, bases de données normatives**. Additif cumulatif, n° 2. Version française. Paris: La Defense, 1987.

BALTZ, C.F. **The need for a thesaurus in automated information retrieval.** New York: Owego; IBM, 1962.

BARHYDT, G.T. & SCHMIDT, C.T. **Information Retrieval Thesaurus of Education Terms.** Cleveland (Ohio): Case Western Reserve University, 1968.

**BASE de datos Thesauri/ditribuida por ECHO.**

BASILIO, M.B.; BRANCHER, M.; FERRAZ BARBOZA, E.M.; **TESAUROS bibliografía, 1983-1984.** Brasilia: Ibict, 1985.

BECK, H. **Book review of International Classification and Indexing Bibliography. V.1: Classification system and thesauri 1950-1982.** Frankfurt: Indeks Verlag, 1982.

BERNIER, C.L. *Correlative indexes II: Correlative trope indexes.* **American Documentation**, v. 8 (1957) pp. 103-122.

BERNIER, C.L. & HEUMANN, K.F. *Correlative indexes III: semantic relations among sermanterms -the technical thesaurus.* **American Documentation**, v. 8 (1957) pp. 211-220.

BERTRAND-GASTALDY, S. & DAVIDSON, C.H. *Improved design of graphic displays in thesauri - through technology and ergonomics.* **Journal of Documentation**, v. 42, n° 4 (1986) pp. 225-251.

BETZ PARKER, Elisabeth, comp. **LC thesaurus for graphics materials: Topical terms for subject access.** Washington: Library of Congress, 1987.

**BIBLIOGRAFÍA de vocabularios, tesauros, encabezamientos de materia y esquemas de clasificación de ciencias sociales (mono y plurilingües).** /preparado por la biblioteca ASLIB en colaboración con Jean Aitchison y C.G. Allen. Paris: Unesco, 1982.

**BIBLIOGRAPHIE des thesauri et index par matières parus depuis 1960: extraits du bulletin des bibliothèques de France 1969-1974.** /par Marie Thèrese Laureilhe. Paris: Bibliothèque Nationale, 1975.

**BIBLIOGRAPHY of United Nations Thesauri, classifications, nomenclatures.** Genève: Inter-Organization board for information systems, 1979.

BLAGDEM, J.F. Thesaurus compilation methods. A literature review. **Aslib Proceedings**, v. 20, n° 8 (1968) pp. 345-359.

BOISSIÈRE, P. **Dictionnaire analogique de la langue française, Répertoire complet des MOTS par les IDÉES et des IDÉES par les MOTS**. Paris: Aug. Boyer et Ce., Libraires-Éditeurs, 1862.

BOLL, J.J. From subject headings to descriptors: the hidden trend in Library of Congress Subject Headings. **Cataloguing and Classification Quarterly**, v. 1 (1982) pp. 3-28.

BRITISH STANDARD INSTITUTION (BSI). **BSI Root thesaurus**. 2 v. London: BSI, 1981.

BROWNSON, H. **Proceedings of the international study conference on classification for information retrieval**. Aslib, 1957, pp. 99-100.

BUCHANAN, B. **A glossary of indexing terms**. London: Clive Bingley, 1976.

BUREAU NATIONAL D'INFORMATION SCIENTIFIQUE ET TECHNIQUE. **Annuaire des thésaurus suivi d'un index des domaines et sous-domaines couverts par les thésaurus et d'une Liste des Organismes auteurs**. 3<sup>a</sup> ed. Paris: Comité de Coordination des Thésaurus Sectoriels, 1975.

CABRE CASTELLVI, T. **La terminología: Teoría, metodología, aplicaciones**. Barcelona: Edit. Empúries, 1883.

CURRÁS, E. Sistemas mecanizados de almacenamiento y recuperación de lenguajes gráficos y científicos. Su aplicación en los países de habla hispana. **Revista Española de Documentación Científica**, v. 4, n° 2 (1981) pp. 91-110.

----- **Tratado sobre Ciencia de la Información**. Rosario (Argentina): Universidad Nacional, 1996.

CASARES, J. Nuevo concepto del Diccionario de la Lengua, y otros problemas de Lexicografía y Gramática. Tomo V de las **Obras completas**. Madrid: Espasa Calpe, 1941.

CASARES, J. **Diccionario ideológico de la lengua española**. Barcelona: Gustavo Gili, 1984.

**CATÁLOGO de recursos terminológicos detectados en España.** Antony (Francia): Unión Latina/ICYT, 1995.

CENTRO DE ESTUDIOS Y EXPERIMENTACIÓN DE OBRAS PÚBLICAS. Departamento de Información y Documentación. **Tesouro DIC.** Madrid: Centro de Estudios y experimentación de Obras Públicas, 1982.

COMMISSION DES COMMUNAUTES EUROPEENES. **ASTUTE. Automated System for theaurus updating, testing and editing.** Bruxelles: CEE, 1976.

COMMISSION OF THE EUROPEAN COMMUNITIES. **Thesaurus guide. Analytical directory of selected vocabularies for information retrieval.** Amsterdam: North-Holland, 1979.

COSTELLO, J.C. Uniterm indexing: principles, problems and solutions. **American Documentation**, v. 12, n° 1 (1961) pp. 20-26.

----- **Coordinate indexing.** New Brunswick: Graduate School of Library Service, 1966.

COVARRUBIAS OROZCO, S. de. **Tesoro de la lengua castellana o española.** Madrid: Por Luis Sánchez, 1611.

CHAN, L.M. & POLLARD, R. **Thesauri used in online databases: An Analytical Guide.** New York, etc: Greenwood Press, 1988.

DAHLBERG, I. **Reference Tools and Conferences in Classification and Indexing, International Classification and Indexing Bibliography.** V. 2. Frankfurt: Indeks Verlag, 1984.

----- The International Classification and Indexing Bibliography (ICIB) and its classification systems. **International Classification** v. 12, n° 3 (1985) pp. 143-152.

**DICCIONARIO de la lengua castellana, en que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las frases o modos de hablar, los proverbios o refranes, y otras cosas convenientes al uso de la lengua.** Madrid: En la Imprenta de Francisco del Hierro, Impresor de la Real Academia Española, 1739.

DOGGEN, J. e ILJON, A. **ASTUTE. Automated system for thesaurus updating, testing and editing.** Luxemburgo: Comisión de las Comunidades Europeas, 1976.

**ENCICLOPEDIA universal ilustrada europeo americana.** Madrid: Espasa Calpe, 1928.

**ENGINEERS JOINT COUNCIL, NEW YORK. First Revision of the Thesaurus of Astia Descriptors.** Final report on Contract AF 18(600)1944: Office of Technical Services AD-278, 168, E. Wall, Arlington, Va.: Astia, 1962.

**ESPAÑA.** Ministerio de Educación y Ciencia. **LEDA Thesaurus** [inédito], 1982.

**FEDERATION INTERNATIONALE DE DOCUMENTATION (FID). Bibliography of thesauri and controlled vocabularies.** La Haya: FID, 1977.

**FEDERATION INTERNATIONALE DE DOCUMENTATION (FID). Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües.** Bogotá: FID/CLA, 1980.

**FELBER, H; UNESCO. Terminology manual.** Paris: Unesco, 1984.

**FIELD, B.J.; INSTITUTION OF ELECTRICAL ENGINEERS. Semi-automatic development of thesauri using free-language vocabulary analysis.** London: The Institution of Electrical Engineers, 1975.

**FUNDACION DE LA RED DE INFORMACION CIENTIFICA (FUINCA). Catálogo de bases de datos españolas.** Madrid: FUINCA, 1985.

**FURTH, S.E. Luhn, Hans Peter.** En: **ALA World Encyclopedia of Library and Information Services.** 2nd. ed. Chicago: ALA, 1983.

**GALINDO, F.** Thesaurus jurídico inteligente. En **Actas de las Segundas Jornadas Españolas de Documentación Automatizada.** Málaga: Consejo Superior de Investigaciones Científicas, 1986, pp. 471-482.

**GARCÍA GUTIÉRREZ, A.** Los lenguajes documentales. En: **Fundamentos de información y documentación,** comp. J. López Yepes. Madrid: Eudema, 1990.

**GAUDENZI, N.** Thesaurus : an ancient term for a modern documentation tool. **Bolletino Inf. Doc.,** v. 13, nº 1/4 (1974) pp. 13-22.

**GILBERT, V.** A list of thesauri and subject headings held in the ASLIB library. **ASLIB Proceedings,** v. 31, nº 6 (1979) pp. 264-274.

GORKOVA, V.I.; INTERNATIONAL FEDERATION OF DOCUMENTATION (FID/DT); ALLUNIONSINSTITUT FÜR WISSENSCHAFTLICHE UND TECHNISCHE INFORMATION (VINITI). **Essential problems in terminology for Informatics and Documentation. Frequency terms lists.** Moskva: VINITI, 1982.

GUINCHAT & SKOURI. **Guide pratique des techniques documentaires.** Paris: EDICEF, 1989.

HOLM, B.E. y RAMUSSEN, L.E. Development of a technical thesaurus. *American Documentation*, v. 12, July (1961) pp. 184-190.

INDEKS GMBH; FÉDÉRATION INTERNATIONALE DE DOCUMENTATION (FID). **Classification systems and thesauri 1950-1982.** Frankfurt: Indeks Verlag, 1982.

INFOTERM, International Information Center for Terminology. **Terminologies for the Eighties.** München: K.G. Saur Verlag, 1984.

INFOTERM; KROMMER-BENZ, M. **Networking in terminology. International Cooperation in terminology work. Proceedings of the second Infoterm Symposium, Vienna, 14-17 April 1985.** München: K.G. Saur Verlag, 1986.

INSTITUTE FOR SCIENTIFIC, TECHNICAL AND ECONOMIC INFORMATION (IINTE). **Bibliographic Bulletin of the Clearinghouse IINTE.** Supplement 1980. Warszawa: IINTE, 1981.

INSTITUTO DE CIENCIA Y TECNOLOGIA (ICYT). **Tesaurus SPINES 2 v.** (Versión provisional, sujeto a revisión). Madrid: Consejo Superior de Investigaciones Científicas, 1985.

INSTITUTO EDUARDO TORROJA DE LA CONSTRUCCIÓN Y EL CEMENTO. **Léxico de la construcción.** Madrid: I.E.T.C.y C., 1968.

INTERNATIONAL BUSINESS MACHINES CORPORATION. **Technical Information Center Thesaurus.** New York: IBM, 1964.

**INVENTARIO de lenguajes documentarios no tradicionales en lengua castellana y/o portuguesa/.** Grupo de trabajo sobre lenguajes documentarios Reuniber 78. Buenos Aires: (impreso en el ICYT, Madrid), 1981.

IRAZAZABAL NERPELL, A. de. La terminología ¿por qué?. En **II Encuentro Hispano-Luso de Investigación Científica y Técnica**. Marzo 1988, pp. 323-330.

----- Aplicación de la cadena de elaboración y desarrollo de tesauros en un microordenador a la construcción del tesoro de la programación (1987-1988) del CSIC. En: **Terceras Jornadas Españolas de Documentación Automatizada**. Palma de Mallorca: Consejo Superior de Investigaciones Científicas, 1990, pp. 876-889.

----- La terminología científica. Su enseñanza en lengua española. **Política científica**, v. 38, nov. (1993) pp. 52-58.

IRAZAZABAL NERPELL, A. de; ALVAREZ S., ORTIZ, M.C. y ZARCO, J. Inventario automatizado de recursos terminológicos en lengua española y portuguesa. En **Actas de las Segundas Jornadas Españolas de Documentación Automatizada**. Málaga: Consejo Superior de Investigaciones Científicas . Málaga: Consejo Superior de Investigaciones Científicas, 1986, pp. 111-124.

IRAZAZABAL NERPELL, A. de., ALVAREZ S., ORTIZ, M.C. y ZARCO, J. **Catálogo de recursos terminológicos en lengua española**. Madrid: ICYT, 1987.

IZQUIERDO ARROYO, J.M. & MORENO FERNÁNDEZ, L.M. Listas de encabezamientos de materia y Thesauri en perspectiva comparada. **Documentación de las Ciencias de la Información**, nº 17 (1994).

JONES, K.S. Compound words: a problem in postcoordinate retrieval systems. **Journal of the American Society for Information Science**, v. 22, nº 4 (1971) pp. 242-250.

----- **Automatic keyword classification for information retrieval**. London: Butterwords, 1971.

----- Some thesauric history. **ASLIB Proceedings**, v. 24, Jul. (1972) pp. 400-411.

----- **Linguistics and information science**. New York and London: Academic Press, 1973.

JOYCE & NEEDHAM. The thesaurus approach to information retrieval. **American Documentation** nº 9 (1958) pp. 192-197.

KAZLAUSKAS, E. y HOLTZ, T. The application of a minicomputer to thesaurus construction. **Journal of the American Society for Information Science**, v. 31, n° 5 (1980) pp. 363-368.

KLEINBERT, P. Prolegomenon to "intelligent" thesaurus software. **Journal of Information Science**, v. 11, n° 2 (1985) pp. 45-53.

LAFUENTE CASTAÑEIRA, I. **El concepto de tesoro en España ayer y hoy: 1970-94**. Proyecto de investigación dirigido por la Dra. Emilia Currás. Curso de Especialista Universitario en Documentación Pedagógica 1994/95. Facultad de Ciencias de la Educación. Universidad Complutense de Madrid.

LAGUNA SERRANO, E.; IRAZAZABAL NERPELL, A. de y VALLE BRACERO, A. Confección automática de tesauros. **Revista Española de Documentación Científica**, v. 12, n° 2 (1989) pp. 129-140.

LANCASTER, F.W. Vocabulary control in information retrieval systems. **Advances in librarianship**, n° 7 (1977).

LASSO DE LA VEGA JIMÉNEZ-PLACER. Organización de la Documentación por palabras y descriptores específicos de materia: Los Thesaurus. **Racionalización**, n° 6 (1966) pp. 433-445.

LAUREILHE, M.T. Bibliographie courante des Thesauri et index par matières. **Bulletin des Bibliothèques de France**, n° 3 (1982).

LAUREILHE, M.T. Essai de bibliographie des thésauri et index par matières parus depuis 1960 (jusqu'à Mars 1969). **Bulletin des Bibliothèques de France**, v. 14, n° 5 (1969) pp. 203-218.

LUHN, H.P. A new method for recording and searching information. **Documentation**, v. 4, n° 1 (1953) pp. 14-16.

----- **A Statistical Approach to Mechanised Literature Searching**. New York: IBM Research Center, 1957.

----- **Keyword in Context Index for technical literature**. New York: IBM, 1959.

----- Auto-encoding of documents for information retrieval systems. **Modern trends in documentation**. Pergamon Press, 1959, pp. 45-58.

MARCUS, R.S. An automated expert system for information retrieval. **Proceedings of the 44th Annual ASIS Meeting 18** (1981) pp. 270-273.

MIYAJIMA, T. A study of specialized terminology: the problem of technical terms. **International Classification**, v. 9, n° 3 (1982) pp. 152-154.

MOOERS, C. Zatocoding and developments in Information Retrieval. **Aslib Proceedings**, v. 8 (1956) pp. 1-20.

MOSQUERA MARTINEZ, R.; RODRIGUEZ MARTIN, F. y MOLOWNY GOMEZ, I. Tesoro de drogodependencias. **Cuadernos Técnicos de Estudios y Documentación**. Madrid: Ministerio de Sanidad y Consumo. Monografía n° 11 de Comunidad y drogas, Junio 1990.

NACIONES UNIDAS. **UNBIS thesaurus**. Edición francesa, lista trilingüe. New York: Naciones Unidas, 1986.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA). **NASA thesaurus: subject terms for indexing scientific and technical information**. 3 v. Washington: NASA, 1967.

NEDOBITY, W. Terminology and artificial intelligence. **International Classification**, 12, n° 1 (1985) pp. 17-19.

OFICINA INTERNACIONAL DEL TRABAJO (OIT). **Tesoro OIT. Terminología de trabajo, empleo y formación**. Multilingüe (inglés, francés, español). 2ª ed. Ginebra: OIT, 1978.

OKLA UNIVERSITY, TULSA. **Exploration and production thesaurus; listing of scientific and engineering terms and their relationship for use in vocabulary control in indexing and retrieving Petroleum Abstracts**. Tulsa: Okla University, 1965.

OLMEDO, M.I. [et al]. Bibliografía comentada sobre lenguajes de indización y automatización. **Revista Española de Documentación Científica**, 10, 4 (1987) pp. 453-471.

ORGANIZACION PARA LA COOPERACION Y EL DESARROLLO ECONOMICO (OCDE). **Macrothesaurus: A basis list of Economic and Social Development terms**. Paris: OCDE, 1972.

ORGANIZACION PARA LA COOPERACION Y EL DESARROLLO ECONOMICO (OCDE). **International road research documentation thesaurus (IRRD)**. Paris: OCDE, 1972.

PÉREZ ALVAREZ-OSSORIO, J.R.; RIUDAVETS MONTES, A.; VALLE BRACERO, A. Cambio automático de lenguaje pivote en un tesoro multilingüe informatizado. En: **Actas de las Primeras Jornadas Españolas de Documentación Automatizada**. Madrid: 20-21 Nov. 1984. Madrid: Consejo Superior de Investigaciones Científicas, 1984, pp.601-622.

PULP AND PAPER RESEARCH INSTITUTE OF CANADA. **Thesaurus of pulp and paper terms; a list of keywords with their cross-relationships for use in indexing and retrieving information on pulp and paper technology**. Point Claire, Quebec: Pulp and Paper Research Institute of Canada, 1965.

RIZZO, S. **Bibliografia internazionale di thesauri/International bibliography of thesauri**. Roma: Camera dei Deputati, 1987.

ROBERTS, N. The prehistory of the information retrieval thesaurus. **Journal of Documentation**, v. 40, n° 4 (1984) pp. 271-285.

ROGET, P.M. **Roget's International Thesaurus**. 4<sup>a</sup> ed. revised by Robert L. Chapman. New York, etc.: Harper Collins Publishers, 1977.

SANCHO LOZANO, R. Los tesauros de términos científicos y técnicos en general, y especialmente los relativos a la química (revisión bibliográfica). **Revista Española de Documentación Científica**, v. 8, n° 1 (1985) pp. 17-55.

SCHULTZ, C.K., ed. H.P. **Luhn: Pioneer in Information Science**. New York: Spartan Books, 1968.

SVENONIUS, E. Design of controlled vocabularies in the context of emerging technologies. **Library Science with a slant to Documentation**, v. 25, n° 4, Dec (1988) pp. 215-227.

TAUBE, M. Storage and retrieval of information by means of association of ideas. **American Documentation**, n° 6 (1955).

----- Emerging solutions for storage and retrieval of information. **Documentation Inc.**, 1959, pp. 112-140.

**THESAURI and thesauri construction** /compiled by Mexine Mac Cafferty. London: ASLIB, 1977.

**THESAURI relação bibliografica** /Centro de Documentação científica ultramarina/Junta de Investigações de ultramar. Lisboa: Centro de Documentação Científica, 1971-1977.

**THESAURUS guide: Analytical directory of selected vocabularies for information retrieval/ prepared by Gesells Chaft für information und dokumentation for the Commission of the European Communities.** Amsterdam; North Holland; Luxembourg: Office for Official Publications of the European Communities, 1985.

**THESAURUS of Engineers Terms.** New York: Engineers Joint Council, 1961.

**THESAURUS Textil Internacional.** Tarrasa: Escuela Técnica Superior de Ingenieros Industriales, 1971.

**TINKER, L. An annotated Bibliography of Library Automation 1968-1972.** London: Aslib, 1973.

**UNESCO. Bibliografía de vocabularios, tesauros, encabezamientos de materias y esquemas de clasificación de Ciencias Sociales (Mono y Plurilingües).** Paris: UNESCO, 1982.

**UNESCO. SPINES thesaurus: a controlled and structured vocabulary of science and technology for policy-making, management and development, 3 v.** Paris: UNESCO, 1976.

**U.S. DEFENSE DOCUMENTATION CENTER. ASTIA chemical thesaurus.** Springfield, Virginia: U.S. Defense Documentation Center, 1962.

----- **Thesaurus of ASTIA descriptors.** Springfield, Virginia: U.S. Defense Documentation Center, 1964.

**U.S. RECLAMATION SERVICE. Thesaurus of descriptors; a list of keywords and cross-references for indexing and retrieving the literature of water resources development.** Denver: U.S. Reclamation Service, 1964.

**VICKERY, B.C. Thesaurus: a new word in documentation. Journal of Documentation,** v. 16, n° 4 (1966) pp. 181-189.

WALL, E. A practical system for documenting building research. **Documentation of building science literature**, pub. 791 (1960) p. 15: US Senate Report, 86th Congress, 2nd Session, Document nº 113, 1960.

WALL, E. Symbiotic development of thesauri and information systems: A case history. **Journal of the ASIS**, v. 26, March-April (1975) pp. 71-79.

WELLISCH, H.H. *Indexing and abstracting: a current awareness bibliography*. **Indexer**, v. 15, nº 1, pp. 29-36.

#### **Capítulo 4. La recuperación de información**

ADAMS, R. **Comunicaciones y acceso a la información en la biblioteca**. Madrid: Fundación Germán Sánchez Ruipérez/Pirámide, 1994.

AMAT NOGUERA, N. **La documentación y sus tecnologías**, 2ª ed. Madrid: Pirámide, 1995.

ARENAS ALEGRÍA, L. **Efectividad y dinamismo en la recuperación documental mediante análisis cluster**. Tesis doctoral, Universidad de Deusto, 1991.

BARES, M. **La búsqueda documental en el contexto telemático: Modalidades de automatización y utilización de bases de datos**. Madrid: Díaz de Santos, 1988.

BOOKSTEIN, A. **Implication of boolean structures for probabilistic retrieval: research and development in information retrieval**. Montreal: Association for computing machinery, 1985.

CAREY [et al]. **Manifiesto del sistema de base de datos de tercera generación**. Comité para la Función Avanzada en SGBD. **Novática**, v. 17, nº 91 (1991).

COOPER, W.S. A definition of relevance for information retrieval. **Information storage and retrieval**, v. 7, nº 1 (1971) pp. 19-37.

----- Getting beyond Boole. **Information processing and management**, v. 24, nº 3 (1988) pp. 243-248.

CORNELIO, E.R. **Bases de datos relacionales**. Madrid: Paraninfo, 1988.

CHAUMIER, J. **L'accès automatisé à l'information**. Paris: Entreprise Moderne d'Édition, 1982.

CHEN, C.C. & SCHWEIZER, S. **Online bibliographic searching: a learning manual**. New York: Neal-Schuman, 1981.

DATE, C.J. **An introduction to database systems**. v. I. 5<sup>a</sup> ed. Reading, Massachusetts: Addison-Wesley, 1990.

DEEN, S.M. **Fundamentos de los sistemas de bases de datos**. Barcelona: Gustavo Gili, 1987.

DEWEZE, A. **L'accès en ligne aux bases documentaires**. Paris: Masson, 1983.

DUBOIS, C.P. The use of thesauri in online retrieval. **Journal of Information Science**, v. 8, n<sup>o</sup> 2 (1984) pp. 63-66.

FARRADANE, J.E.L. Analysis and organization of knowledge for retrieval. **ASLIB Proceedings**, v. 22 (1970) pp. 607-616.

FEDERATION INTERNATIONALE DE DOCUMENTATION (FID). **The role of graphic display of concepts relationships in indexing and retrieval vocabularies including a thesaurus of documentation terms**. Bruselas: European Atomic Energy Community (EURATOM), 1965.

FERNÁNDEZ, C. **El modelo relacional de datos: De los fundamentos a los modelos deductivos**. Madrid: Díaz de Santos, 1987.

FOSKETT, A.C. **The subject approach to information**. 5<sup>a</sup> ed. London: Clive Bingley, 1996.

GARCÍA GUTIÉRREZ, A. **Procedimientos de Análisis Documental Automático: Estudio de caso**. Sevilla: Instituto Andaluz del Patrimonio Histórico, 1996.

GARCÍA MARCO, F.J. De la consulta de catálogos a la gestión de información: Tensiones hacia el cambio en el diseño de OPACs. **Boletín de la ANABAD**, v. 41, n<sup>o</sup> 3-4 (1991).

----- Interfaces amigables para la recuperación de la información bibliográfica. **Scire**, v. 1, n° 1 (1995).

GARDARIN, G. **Bases de datos**. Madrid: Paraninfo, 1987.

GARDIN, J.C. **Systèmes experts et Sciences humaines**. Paris: Eyrolles, 1987.

GILCHRIST, A. **The thesaurus in retrieval**. London: Aslib, 1971.

HARTLEY, R.J. **Online searching: principles & practice**. London, etc.: Bowker-Saur, 1990.

HARTNER, E.P. **An introduction to automated literature searching**. New York: Dekker, 1981.

HAWKINS, D.T. Applications of artificial intelligence (AI) and expert systems for online searching. **Online**, v. 12, n° 1 (1988) pp. 31-40.

HOOVER, R.E. **Online search strategies**. White Plains, NY.: Knowledge Industry, 1982.

KEEN, E.M. In the processing of printed subject entries during searching. **Journal of Documentation**, v. 33, 4 (1977) pp. 266-276.

----- Some aspects of proximity searching in text retrieval systems. **Journal of Information Science**, n° 18 (1992) pp. 89-98.

KEHOE, C.A. Interfaces and expert systems for online retrieval. **Online**, v. 9, n° 6 (1985) pp. 489-505.

KORTH, H.F. & SILBERSCHATZ, A. **Database system concepts**. New York: McGraw-Hill, 1986.

LÓPEZ RINCÓN, C. **Diseño e implementación de un sistema automático de análisis documental**. Madrid: Universidad Complutense, Dept. de Lógica y Filosofía de la Ciencia, 1995. (Tesis doctoral).

LANCASTER, F.W. **Vocabulary control for information retrieval**. Washington, D.C.: Information Resources Press, 1972.

LUHN, H.P. A statistical approach to mechanized encoding and searching of literary information. **IBM Journal of research and development**, v. 1, n° 4, October (1957) pp. 309-317.

----- A business intelligence system. **IBM Journal of research and development**, n° 2 (1985) pp. 314-319.

MANIEZ, J. **Le rôle de la syntaxe dans les systèmes de recherche documentaire**. Dijon: Institut Universitaire de Technologie, 1976-1977, 2 v.

MARCUS, R.S. An experimental comparison of the effectiveness of computers and humans as search intermediaries. **Journal of the American Society for Information Science**, v. 34, 6 (1983) pp. 381-404.

MARON, M.E. & KUHNS, J.L. On relevance probabilistic indexing and information retrieval. **Journal of the ACM**, v. 7, n° 3 (1960).

MATTHEWS, J.R. **Public Access to Online Catalogs**. 2<sup>a</sup> ed. New York: Neal Schuman, 1985.

MIGUEL CASTAÑO, A. de & PIATTINI VELTHUIS, M.G. **Concepción y diseño de bases de datos: Del modelo E/R al modelo relacional**. Madrid: RA-MA, 1993.

MOYA ANEGÓN, F. de. **Los sistemas integrados de gestión bibliotecaria**. Madrid: Anabad, 1995.

ODDY, R.N. [et al]. Information retrieval through man-machine dialogue. **Journal of Documentation**, v. 33, 1 March (1977) pp. 1-14.

----- **Information retrieval research**. London: Butterworths, 1981.

PRASSE, M.J. Interface design procedure. **Annual review of OCLC research**. July (1990) June (1991), pp. 17-18.

ROBERTSON, S.E. Theories and models in information retrieval. **Journal of Documentation**, v. 33, n° 2, June (1977) pp. 126-148.

----- The probability ranking principle in IR. **Journal of Documentation**, v. 33, n° 4, December (1977) pp. 294-304.

----- Probability of relevance: a unification of two competing models for documental retrieval. **Information technology research and development**, v. 1, n° 1 (1982) pp. 1-21.

ROBSON, A. & LONGMAN, J.S. Automatic aids to profiles construction. **Journal of ASIS**, v. 27, n° 4, July-August (1976) pp. 213-223.

SALTON, G. A simple blueprint for automatic boolean query processing. **Information processing and management**, v. 24, n° 3 (1988) pp. 269-280.

SCHANK, R. **El ordenador inteligente**. Barcelona: A. Bosch, 1986.

SCHAUB, J.A. **CD-Rom for public access catalogs**. *Library High Tech*, v. 3 n° 3 (1985) pp. 7-13.

SERVAN-SCHREIBER, J.J. **El desafío mundial**. Madrid: Plaza & Janes, 1980.

STANDERA, O. On-line retrieval systems: some observation on the user/system interface. En: **Proceedings of the ASIS Annual Meeting**, v. 12: **Information revolution**. Washington: ASIS, 1975, pp. 38-40.

STONEBRAKER, M. **The INGRES papers: The anatomy of a relational database management system**. Reading, Massachusetts: Addison Wesley, 1986.

VAN RIJSBERGEN, C.J. **Information retrieval**. London: Butterworths, 1979.

WALKER, S. Improving subject access painlessly: recent work on the OKAPI online catalogue projects. **Program**, v. 22 n° 1 (1988) pp. 21-31.

WIEDERHOLD, G. **Diseño de bases de datos**. México: McGraw-Hill, 1985.

**Capítulo 5. Análisis semántico y estructural de diferentes tesauros y de las bases de datos en las que se utilizan.**

AITCHISON, T.M., A.M. HALL, and J.M. TRACY. **Comparative Evaluation of Index Languages. Part II. Results.** London: Institution of Electrical Engineers, 1970.

ALVARO BERMEJO, C.; VILLAGRA RUBIO, A. y SORLI ROJO, A. Desarrollo de lenguajes documentales formalizados en lengua española: Una evaluación. Parte I: Vigencia teórica y práctica de lenguajes controlados. **Revista Española de Documentación Científica**, v. 12, nº 2 (1989) pp. 154-159.

ALVARO BERMEJO, C., VILLAGRA RUBIO, A. y SORLI ROJO, A. Desarrollo de lenguajes documentales formalizados en lengua española. Parte II: Evaluación de los tesauros disponibles en lengua española. **Revista Española de Documentación Científica**. v. 12, nº 3 (1989) pp. 283-305.

BUREAU MARCEL VAN DIJK. **Definition of Thesauri Essential Characteristics.** 2 vols. Brussels: Bureau M. Van Dijk, 1976.

DAHLBERG, I. "The Broad System of Ordering (BSO) as a Basis for an Integrated Social Sciences Thesaurus?". **International Classification**, v. 7, nº 2 (1980) pp. 66-72.

GERD, A.S. Associative Relations in an Information Retrieval Thesaurus and Ways to Identify Them. **Automatic Documentation and Mathematical Linguistics**, v. 14 nº 3 (1980) pp. 25-30.

JONES, K.P. Problems Associated With the Use of Compound Words in Thesauri With Special Reference to BS 5723: 1979. **Journal of Documentation**, v. 37, nº 2, June (1981) pp. 53-68.

KOCHEN, M. & TAGLIACOZZO, R. A Study of Cross-Referencing. **Journal of Documentation**, v. 24, nº 3 (1968) pp. 173-191.

MAGARIÑOS, A. [et al]. El tesoro de Medio Ambiente de la Secretaría General de Medio Ambiente. **Boletín de la ANABAD**, v. 4, oct.-dic. (1990) pp. 141-162.

**ROOT Thesaurus.** London: British Standard Institution, 1981.

SHUMWAY, R.H. Some Estimation Problems Associated with Evaluating Information Retrieval Systems. En: **Evaluation of document Retrieval Systems: Literature Perspective, Measurement, Technical Papers**. Bethesda, Md.: Westat Research, Inc., 1968, pp. 78-96.

SOERGEL, D. The Broad System of Ordering: A Critique. **International Forum on Information and Documentation**, v. 4, n° 3, July (1979) pp. 21-24.

**SPINES Thesaurus**. Paris: Unesco, 1976.

WELLISCH, H.H. The Broad System of Ordering or Bishop Wilkins Redivivus: A Review Article. **Library Quarterly**, v. 49, n° 4, October (1979) pp. 444-452.

### **Capítulo 6. Evaluación del rendimiento de tesauros españoles en sistemas de información.**

AITCHISON, J. et al. **Comparative evaluation of index languages**. London: Institution of Electrical Engineers, 1970.

ANGUERA, M.T. **Metodología de la observación en las Ciencias humanas**. Madrid: Cátedra, 1989.

BARKER, F.H.; VEAL, D.C.; WYATT, B.K. Comparative efficiency of searching titles, abstracts and index terms in a freetext data base. **Journal of Documentation**, v. 28, n° 1 (1972) pp. 22-36.

BATES, M.J. How to use controlled vocabularies more effectively in online searching. **Online**, v. 12, n° 6 (1988) pp. 45-56.

BETTS, R. & MARRABLE, D. Free text vs controlled vocabulary: retrieval precision and recall over large databases. En: **Online Information 91**. Proceedings of the 15th International Online Information Meeting, London, 10-12 December 1991 Oxford New Jersey: Learned Information (Europe) Ltd, 1991, pp. 153-165.

BHATTACHARYYA, K. The effectiveness of natural language in science indexing and retrieval. **Journal of Documentation**, v. 30, n° 3 (1974) pp. 235-254.

BIVINS, K.T. **New trends in documentation and information**. London: ASLIB, 1980.

BLAIR, D.C. **Language and representation in information retrieval**. Amsterdam, etc.: Elsevier, 1990.

BLAIR, D.C. & MARON, M. E. An evaluation of retrieval effectiveness for a full text document retrieval system. **Communications of the ACM**, v. 28, n° 3 (1985) pp. 289-299.

BOURNE, C.P. Evaluation of indexing systems. **Annual review of information science and technology**, n° 1 (1966).

BRIER, S. Cybersemiotics: A new interdisciplinary development applied to the problems of knowledge organisation and document retrieval in information science. **Journal of Documentation**, v. 52, n° 3 (1996) pp. 296-344.

BROOKS, H.M. Experts systems and intelligent information retrieval. **Information processing and management**, v. 23, n° 4 (1987) pp. 367-382.

CALKINS, M.L. Free text or controlled vocabulary?. **Database** v. 3, n° 2, June (1980) pp. 53-67.

CLEVERDON, C.W. **ASLIB Cranfield research project: report on the testing and analysis of an investigation into the comparative efficiency of indexing systems**. Cranfield: College of Aeronautics, 1962.

----- On the inverse relationship of recall and precision. **Journal of documentation**, v. 28, n° 3 (1972) pp. 195-201.

----- Evaluation tests of information retrieval systems. **Journal of Documentation**, v. 26, n° 1 (1970) pp. 55-67.

----- **The Methodology of Evaluation of Operational Information Retrieval Systems Based on the Medlars**. Cranfield: Cranfield Institute of Technology Report, 1968.

CLEVERDON, C.W. & THORNE, R.G. **A brief Experiment with the Uniterm System of Coordinate Indexing for the Cataloging of Structural Data**. RAE Library Memorandum N° 7, AD 35004. Farnborough, England: Royal Aircraft Establishment, 1954.

CLEVERDON, C.W. & MILLS, J. The testing of index language devices. **Aslib Proceedings**, v. 15, n° 4 (1963) pp. 106-130.

CLEVERDON, C.W.; MILLS, J.; KEEN, M. **Factors determining the performance of indexing systems**. 2 v. Cranfield: The National Science Foundation; ASLIB, 1966.

COOPER, W.S. A definition of relevance for information retrieval. **Information storage and retrieval**, v. 7 n° 1 (1971) pp. 19-37.

CROFT, W.B. A model of cluster searching based on classification. **Information Systems**, n° 5 (1980).

DIMITROFF, A. & WOLFRAM, D. The effect of linkage structure on retrieval performance and user response in a hypertext-based bibliographic retrieval system. **Ann. Rev. OCLC Research** 1995, pp. 58-60.

DRINKWATER, C. Social and behavioural sciences. En: **Manual of Online Search Strategies**, ed. por C.J. Armstrong & J.A. Large. Aldershot: Gower Press, 1988, pp. 469-506.

----- Free text vs. controlled vocabulary: A reassessment. **Online Review**, v. 11, n° 10 (1987) pp. 243-253.

ELLIS, D. **New horizons in information retrieval**. London: The Library Association, 1990.

FAIRTHORNE, R. Automatic retrieval of recorded information. **Computer journal** (1958).

FESTINGER, L.; KATZ, D. (eds.). **Los métodos de investigación en las Ciencias sociales**. Buenos Aires: Paidós, 1972.

FIDEL, R. Searchers selection of search keys: 2. Controlled vocabulary or free-text searching. **Journal of the American Society for Information Science**. v. 42, n° 7, Aug (1991) pp. 501-514.

FOSKETT, A.C. The evaluation of information retrieval systems. En: **The subject approach to information**. London: Clive Bingley, 1982, pp. 518-551.

**FREE text retrieval systems: A review and evaluation**. London: Taylor Graham, 1989.

GIERE, W. & DETTMER, H. Free-text classification and retrieval based on a thesaurus. Eight years of experience at the Johann Wolfgang-Goethe University, Medical School. En: **10th. Annual Symp. Comput. Appl. Med. Care**. Washington: (s.n.), 1986, pp. 85-88.

GILCHRIST, A. **The thesaurus in retrieval**. London: ASLIB: 1971.

GOFFMAN, W. & NEWILL, V.A. A methodology for test and evaluation of information retrieval systems. **Information storage and retrieval**, n° 3, (1966) pp. 19-25.

GULL, D. Seven years of work on the organisation of materials in the special library. **American Documentation**, v. 7 (1956).

HARTER, S.P. **Online information retrieval: Concepts, principles and techniques**. Orlando, Florida: Academic Press, 1986.

HARTER, S.P. & CHENG, Y. Colinked descriptors: Improving vocabulary selection for end user searching. **Journal of the American Society for Information Science**, v. 47, n° 4 (1996) pp. 311-325.

HENDERSON, M.M. **Evaluation of information systems: A selected bibliography with informative abstracts**. NBS Technical note 297. Washington, D.C.: National Bureau of Standards, 1967.

HERSEY, D.F. [et al]. Free text word retrieval and scientist indexing: performance profiles and costs. **Journal of Documentation**, v. 27, n° 3 (1971) pp. 167-183.

HIRSCHHEIM, R.A. & SMITHSON, S.C. A critical analysis of information systems evaluation. en Bjorn-Andersen, N. & Davis, G.B. (eds.) **Information Systems Assesment: Issues and Challenges**, Proc. IFIP WG 8.2 Working Conf. Noordwijkerhout, The Netherlands, 27-29 August 1986. North Holland, Amsterdam: IFIP WG, 1986.

JACHNA, J.S.; POWSNER, S.M.; MILLER, P.L. Augmenting GRATEFUL MED with the UMLS Metathesaurus: an initial evaluation. **Bulletin of the Medical Library Association**, v. 81, n° 1, Jan (1993) pp. 20-28.

JONES, K.S. Natural language access to databases: some questions and a specific approach. **Journal of Information Science**, v. 4, n° 1 (1982) pp. 41-48.

JONES, K.S. & RIJSBERGEN, C.J. van. Information retrieval test collections. **Journal of documentation**, v. 32, n° 1 (1976) pp. 59-75.

JONES, S. A thesaurus data model for an intelligent retrieval system. **Journal of Information Science**, v. 19, n° 3 (1993) pp. 167-178.

JONES, K.P. **Prospects for information retrieval**. Informatics 10. Proceedings of a conference jointly sponsored by Aslib Informatics Group, and the Information Retrieval Specialist Group of the British Computer Society. King's College, Cambridge, 21-23 March 1989. Edited by K. P. Jones. Cambridge: Aslib, 1990.

KEEN, M. **On the performance of nine printed subject index types; a selective report of EPSILON**. Aberystwyth: College of Librarianship Wales, 1978.

----- The Aberystwyth index languages test. **Journal of Documentation**, v. 29, n° 1 (1973) pp. 1-35.

KEEN, M. y DIGGER, J.A. (1972). **Report of an Information Science index languages test**. Aberystwyth: College of Librarianship, 1972.

KISH, L. Selección de la muestra. En: L. Festinger & D. Katz, **Los métodos de investigación en las Ciencias sociales**. Barcelona: Paidós, 1953, reimp. 1987.

KRIPPENDORF, K. **Metodología del análisis de contenido**. Barcelona: Paidós, 1990.

KRISTENSEN, J. y JARVELIN, K. The effectiveness of a searching thesaurus in free-text, searching in a full-text database. **International Classification**, v. 17, n° 2 (1990) pp. 77-84.

LANCASTER, F.W. **Guidelines for the evaluation of information systems and services**. Paris: Unesco, 1978.

----- **Information retrieval systems: Characteristics, testing and evaluation**. 2nd. ed. New York: Willey, 1979.

----- **Evaluation of the MEDLARS Demand Search Service**. Bethesda, Maryland: National Library of Medicine, 1968.

----- Precision and recall. En: **Encyclopedia of Library and Information Science**. New York: Marcel Dekker, 1978, v. 23, pp. 170-180.

----- Natural language in information retrieval. En: Lancaster, F.W. **Information retrieval systems, characteristics testing and evaluation**. 2<sup>a</sup> ed. New York: Wiley, 1979. Cap. 20, pp. 279-292.

----- Evaluation and testing of information retrieval systems. En: **Encyclopedia of Library and Information Science**. New York: Marcel Dekker, 1978, v. 8, pp. 234-259.

----- Evaluating the performance of a large computerized information service. **Journal of the American Medical Association**, v. 207, n<sup>o</sup> 1 (1969) pp. 114-120.

LANCASTER, F.W. & GILLESPIE, C.J. Design and evaluation of information systems. **Annual review of information science and technology**, n<sup>o</sup> 5 (1970).

LANCASTER, F.W. & MILLS. Testing indexes and index language devices: The Cranfield Project. **American Documentation**, n<sup>o</sup>. 15 (1964) pp. 4-13.

MALDONADO, A. El control del lenguaje en la producción de una base de datos: el tesoro de urbanismo ISOC. En: **Actas de las Segundas Jornadas Españolas de Documentación Automatizada**. Málaga: Consejo Superior de Investigaciones Científicas, etc., 1986, pp. 125-136.

MALRIEU, D. (1983). La base de donnes ERIC: evaluation de son utilisation et discussion des choix du système. **Brisés**, n<sup>o</sup> 3 (1983) pp. 53-62.

MANIEZ, Jacques. Problèmes de syntaxe dans les systèmes de recherche documentaire. **Documentaliste**, v. 20, n<sup>o</sup> 2 (1983) pp. 52-58.

MARKEY, K.; ATHERTON, P.; NEWTON, C. An analysis of controlled vocabulary and free text search statements in online searches. **Online Review**, v. 4, n<sup>o</sup> 3 (1980) pp. 225-236.

MARTIN, W.A. Methods for evaluating the number of relevant documents. **Journal of the American Society for Information Science**, v. 34, n<sup>o</sup> 6 (1983) pp. 173-177.

McCAIN, K.W.; WHITE, H.D.; GRIFFITH, B.C. Comparing retrieval performance in online data bases. **Information processing and management**, v. 23, n<sup>o</sup> 6 (1987) pp. 539-553.

McJUNKIN, M.C. Precision and recall in title keyword searches. **Information Technology & Libraries**, v. 14, n° 3 (1995) pp. 161-171.

MEYRIAT, J. Social Science information languages: a comparative analysis. **International Classification**, v. 7, n° 2 (1980) pp. 60-65.

MOLL, W.J.D. MEDLINE evaluation study. **Bulletin of the Medical Library Association**, v. 62, n° 1 (1974) pp 1-15.

NUYL, T. W. **Examination of the validity of the conclusions arrived at in the Aslib Cranfield Research Project**. Copenhagen: Danish Center for Documentation. (FID/CR report; 7), 1968.

OLDROYD, B.K. & CITROEN, C.L. Study of strategies used in on-line searching. **Online review**, v. 1, n° 4 (1977) pp. 295-310.

PASQUAL, G. Development of an agricultural database for dissemination of research information to research and extension workers using STATUS software. **Program**, v. 20, n° 3 (1986) pp. 323-331.

PÉREZ, E. Text enhancement, controlled vocabulary versus free text. **Special Libraries**, v. 73, n° 2 (1984) pp. 51-58.

PERRY, J.; KENT, A.; BERRY, M. **Machine literature searching**. New York, etc.: Interscience Publishers, 1956.

RAITT, D.I. Recall and precision in interactive bibliographies search and retrieval systems. **Aslib Proceedings**, v. 36, n° 5 (1984) pp. 229-234.

REGAZZI, J.J. Evaluating indexing systems: a review after Cranfield. **The Indexer**, v. 12, n° 1 (1980) pp. 14-21.

----- Performance measures for information retrieval systems: An experimental approach. **Journal of the American Society for Information Science**, v. 39, n° 4 (1988) pp. 235-251.

**THE RETRIEVAL of information in the humanities and the social sciences: problems as aids to learning**. Ed. T. P. Slavens. New York & Basel: Marcel Dekker, 1981.

REY ROCHA, J. **Tesauro ICYT de biología animal**. Madrid: CINDOC, 1995.

ROBERTSON, S.E. The parametric description of retrieval tests. **Journal of Documentation**, v. 25, n° 1, (1969) pp. 1-27 y v. 25, n° 2, pp. 93-107.

ROTHENBERG, D.H. an efficiency model and a performance function for an information retrieval system. **Information storage and retrieval**, v. 5, n° 3 (1969) pp. 109-122.

ROWLEY, J.E. A comparison between free language and controlled language indexing and searching. **Information Services and Use**, 10 (1990) pp. 147-155.

SALTON, G. A new comparison between conventional indexing (MEDLARS) and automatic text processing (SMART). **Journal of the American Society for Information Science**, v. 23, n° 2 (1972) pp. 75-84.

----- **Automatic information organization and retrieval**. New York, etc.: McGraw Hill, 1968.

----- The evaluation of automatic retrieval procedures selected test results using the SMART system. **American documentation**, v. 16, n° 3 (1965) pp. 209-222.

SALTON, G. y MCGILL, M.J. **Introduction to modern Information Retrieval**. New York: McGraw-Hill, 1983.

SAN MILLAN, M.J. Las bases de datos españolas en Ciencias humanas y sociales: las bases de datos ISOC. En: **Actas de las Segundas Jornadas Españolas de Documentación Automatizada**. Málaga: Consejo Superior de Investigaciones Científicas, etc., 1986, pp. 41-45.

SARACEVIC, T. **An inquiry into testing of information retrieval systems, Part I: Objectives, methodology, design and controls, comparative systems laboratory, Report N°**. CSL: TR-FINAL-1. Cleveland, Ohio: Western Reserve University, 1968.

----- Relevance: a Review and a Framework for the Thinking on the Notion in Information Science. **Journal of the American Society for Information Science**, v. 26, n° 6, Nov.-Dec. (1975) pp. 321-343.

----- Comparative effects of titles, abstracts and full texts on relevance judgments. **Proceedings of the ASIS**, n° 6 (1969) pp. 293-299.

SARACEVIC, T. [et al]. A study of information seeking and retrieving I. Background and methodology; II. Users, questions and effectiveness ; III. Searchers, searches and overlap. **Journal of the American Society for Information Science**, v. 39, n° 3 (1988).

SCHABAS, A.H. Postcoordinate retrieval: a comparison of two indexing languages. **Journal of the American Society for Information Science**, v. 33, n° 1 (1982) pp. 32-37.

SCRIVEN INVERNESS, M. **Evaluation thesaurus**. Pt. Reyes, Ca: Edgepress, 1980.

SHUMWAY, R.H. Some estimation problems associated with evaluating information retrieval systems. En: Westat Research Inc. **Evaluation of Document Retrieval Literature Perspective, Measurement, Technical Papers**. Bethesda, Maryland: Westat Research Inc., 1968, pp. 78-96.

SMITHSON, S. *The evaluation of information retrieval systems: a case study approach*. En: **Prospects for intelligent retrieval. Informatics 10**. Proceedings of a conference jointly sponsored by Aslib Informatics Group, and the Information Retrieval Specialist Group of the British Computer Society. King's College, Cambridge, 21-23 March 1989. London: Aslib, 1990, pp. 75-89.

SNYDER, M.B. [et al]. **Methodology for test and evaluation of document retrieval systems: A critical review and recommendations**. McLean, Virginia: Human Sciences Research Inc., 1966.

SVENONIUS, E. **The effect of indexing specificity on retrieval performance**. Washington: National Science Foundation, 1971.

----- Unanswered questions in the design of controlled vocabularies. **Journal of the American Society for Information Science**, v. 37, n° 5 (1986) pp. 331-340.

SWANSON, D.R. The evidence underlying the Cranfield results. **Library Quarterly**, n° 35 (1965) pp. 1-20.

SWETS, J.A. Effectiveness of information retrieval methods. **American Documentation**, v. 20, n° 1 (1969) pp. 72-89.

TAGUE-SUTCLIFFE, J.M. (ed.). Evaluation of information retrieval systems. Special topic issue. **Journal of the American Society for Information Science** v. 47, n° 1 (1996) pp. 1-105.

**TESAURO Electrotécnico.** Madrid: Asociación de Investigación Industrial Eléctrica (ASINEL), 1993.

**TESAURO de Asuntos sociales.** 2ª ed. Madrid: Ministerio de Asuntos sociales, 1994.

**TESAURO de Medio Ambiente.** 2ª ed. Madrid: Ministerio de Obras Públicas y Urbanismo, 1990.

**TESAURO "Mujer".** Madrid: Instituto de la Mujer, 1993.

**TESAURO ISOC de Psicología.** 2ª ed. Madrid: CINDOC, 1995.

VAN RIJSBERGEN, C.J. [et al]. The selection of good search terms. **Information Processing & Management**, nº 17 (1981) pp. 77-91.

WALLIS, P. & THOM, J.A. Relevance judgements for assessing recall. **Information Process. & Management**, v. 32, nº 3 (1996) pp. 273-286.

WILBUR, W.J. Human subjectivity and performance limits in document retrieval. **Inform. Process. & Management**, v. 32, nº 5 (1996) pp. 515-527.