

Carmen Caro Castro

**EL ACCESO POR MATERIAS EN LOS
CATÁLOGOS EN LÍNEA:
ÍNDICES Y TERMINOLOGÍA DE
LOS USUARIOS EN EL CATÁLOGO CISNE
(UNIVERSIDAD COMPLUTENSE DE MADRID)**



EDICIONES UNIVERSIDAD DE SALAMANCA

COLECCIÓN VITOR

158

c

Ediciones Universidad de Salamanca
y Carmen Caro Castro

1ª edición: Abril, 2005

I.S.B.N.: 84-7800-555-2

Depósito Legal: S. 412-2005

Ediciones Universidad de Salamanca
Plaza de San Benito, s/n
37002 Salamanca (España)
Página en internet: <http://webeus.usal.es>
Correo-e: eus@usal.es

Realización:
Nemática, S.L.

Impreso en España – Printed in Spain

*Todos los derechos reservados.
Ni la totalidad ni parte de este libro
puede reproducirse ni transmitirse
sin permiso escrito de
Ediciones Universidad de Salamanca*

Resumen

El objetivo de este trabajo ha sido estudiar las características de los enunciados de búsqueda empleados por los usuarios de un catálogo público de acceso en línea y comprobar su grado de coincidencia con los índices de encabezamientos de materia.

Se ha realizado un análisis transaccional de 1925 enunciados utilizados en el campo materias en el catálogo CISNE de la Universidad Complutense de Madrid. Para cada enunciado válido se han considerado las siguientes variables: presencia de errores ortográficos y tipográficos, uso de idiomas diferentes al español, número de palabras, tipo de búsqueda, estructura sintáctica y nivel de coincidencia con el índice alfabético de materias. También se ha comprobado si existen diferencias entre las áreas de conocimiento en cuanto a todas las variables analizadas.

Los resultados sobre las características de los enunciados confirman que los usuarios del OPAC mayoritariamente utilizan pocas palabras (1 o 2) para expresar las materias sobre las que necesitan información. Aunque la estructura sintáctica de las expresiones es muy variada, predominan las más simples y en orden directo (*nombre* o *nombre + adjetivo*). El tipo de búsqueda más frecuente es la temática, frente a los escasos porcentajes que se refieren a personas, lugares, entidades o títulos de obras. Los niveles de coincidencia exacta con los encabezamientos de materia del índice son de un 42% mientras que la coincidencia parcial o de algún término alcanza el 47%. Respecto a las diferencias entre las áreas de conocimiento, tanto en las características de los enunciados como en la equiparación, permiten establecer similitudes entre las áreas de “ciencias” y las de “letras”. Sin embargo, sólo dos grupos presentan perfiles claramente diferenciados: Humanidades y Ciencias Médicas.

Abstract

The aim of this work was to study the characteristics of user queries in online catalogs and to check their compatibility with the subject heading index.

A transaction log analysis was performed of 1925 subject queries extracted from the CISNE online catalog at the Universidad Complutense, Madrid. The following variables were considered for every query: occurrence of spelling errors, use of languages different from Spanish, number of words, type of search, syntactic structure and matching between queries and the subject heading index. It was also considered whether there were differences depending on knowledge areas for all the cited variables.

Results for the characteristics of queries show that the majority of OPAC users enter only a few words (1 or 2) in order to express the subject they are seeking. Although the syntactic structures of queries are heterogeneous, the simplest ones and the order *noun* or *noun + adjective* predominate. Topical searching is the most frequent, as opposed to the low percentage of queries for names of persons, geographical and corporate names or titles of works. The percentage for exact matches with subject headings from the index is 42% and the percentage for partial matches or for only some of the query terms is 47%. In relation to the differences between knowledge areas, only two groups differ clearly in their features: the humanities and medical sciences.

Índice

Índice de figuras.....	8
Índice de tablas.....	13
Introducción	15
1. La búsqueda por materias en los catálogos en línea: estado de la cuestión	22
1.1 La búsqueda de información en el OPAC	23
1.1.1 Comportamiento de búsqueda de los usuarios	29
1.1.2 Métodos de búsqueda	35
1.1.3 Problemas de la búsqueda por materias.....	43
1.2 Información temática en los catálogos en línea.....	49
1.2.1 Información temática no normalizada	50
1.2.2 Encabezamientos de materia	53
1.2.2.1 Sintaxis.....	58
1.2.2.2 Estructura semántica	64
1.2.2.3 Registros de autoridad de materia.....	67
1.2.3 Códigos de clasificación.....	70
1.2.3.1 Formato MARC para códigos de clasificación.....	74
1.2.4 Índices y opciones de búsqueda	77
1.2.5 Presentación de los resultados	80

1.3 Mejoras en el diseño de los catálogos en línea.....	84
1.3.1 Estructura de los registros.....	86
1.3.2 Contenido informativo del registro bibliográfico.....	92
1.3.3 Técnicas de recuperación.....	96
1.3.3.1 Simplificación de las técnicas tradicionales	99
1.3.3.2 Técnicas avanzadas de recuperación de Información.....	100
1.3.3.3 Modificación automática de la búsqueda.....	112
1.3.4 Estructuras de conocimiento.....	115
1.3.4.1 Organización de la información.....	121
1.3.4.2 Enriquecimiento de la terminología.....	123
1.3.4.3 Interconexión y compatibilidad	128
1.3.5 La interfaz.....	132
1.3.5.1 Visualización de los índices.....	134
1.3.5.2 Sistemas expertos.....	141
1.3.5.3 Interfaces gráficas	143
1.4 Líneas de investigación	149
1.4.1 Técnicas de recogida de datos y métodos de análisis	157
1.4.2 Variables.....	163
1.4.3 La investigación en España	171
2. Objetivos y Metodología.....	175
2.1 Objetivos e hipótesis de trabajo	180
2.2 Metodología	184
2.2.1 Materiales	185
2.2.1.1 Recogida y depuración de datos.....	185
2.2.1.2 Selección de la muestra	191

2.2.2 Variables.....	193
2.2.2.1 Errores formales.....	193
2.2.2.2 Características de los enunciados	198
2.2.2.3 Comparación con el índice	200
2.2.3 Análisis	212
3. Resultados	215
3.1 Errores formales	216
3.2 Características de los enunciados	222
3.3. Similitud con el índice.....	232
4. Discusión.....	247
4.1 Distribución general	248
4.1.1 Aspectos formales	248
4.1.2 Coincidencia con el índice de encabezamientos de materia	254
4.2 Diferencias entre áreas temáticas	262
5. Conclusiones	269
5.1 Conclusiones generales	270
5.2 Propuestas de mejora.....	280
5.3 Sugerencias para futuras investigaciones.....	290
6. Bibliografía	293
7. Relación de sitios web consultados.....	363
8. Anexo: Tablas y gráficos	366

Índice de figuras

Fig. 1.1	Modelo estratificado de interacción.....	33
Fig. 1.2	Registro bibliográfico. Formato IBERMARC.....	52
Fig. 1.3	Campos IBERMARC para encabezamientos de materia	53
Fig. 1.4	Encabezamientos de materia. Sintaxis. Combinación de conceptos.....	60
Fig. 1.5	Encabezamientos de materia. Estructura semántica	66
Fig. 1.6	Registro de autoridad y referencia comprensiva (IBERMARC)	69
Fig. 1.7	Formato MARC 21. Campos para códigos de clasificación.....	71
Fig. 1.8	Formato MARC 21 para datos de clasificación.....	75
Fig. 1.9	Presentación de resultados. Índice alfabético	81
Fig. 1.10	Presentación de resultados. Registro abreviado.....	82
Fig. 1.11	Técnicas de recuperación de información	97
Fig. 1.12	Visualización de índices. Ordenación de subcampos	137
Fig. 1.13	Visualización de índices. Ordenación estructurada	138
Fig. 1.14	Visualización de índices. Orden cronológico	138
Fig. 1.15	Visualización de índices. Contracción.....	139
Fig. 1.16	Visualización de índices. Relaciones semánticas	139
Fig. 1.17	Interfaces gráficas. Proyecto WEBSOM.....	146
Fig. 1.18	Interfaces gráficas. HyperSpace	147
Fig. 1.19	Interfaces gráficas. Liber World: esferas de relevancia.....	148
Fig. 1.20	Interfaces gráficas. Liber World: conos de navegación.....	149
Fig. 1.21	Técnicas de recogida de datos y métodos de análisis	159
Fig. 1.22	Variables: características de los usuarios.....	164
Fig. 1.23	Variables: características del sistema	165
Fig. 1.24	Variables: entorno de búsqueda.....	167
Fig. 1.25	Variables: proceso de búsqueda.....	158
Fig. 1.26	Variables: rendimiento.....	170

Fig. 2.1 Distribución de las búsquedas por grupos temáticos	190
Fig. 3.1 Enunciados de búsqueda no válidos: tipos	217
Fig. 3.2. Errores tipográficos: tipos.....	217
Fig. 3.3 Enunciados no válidos: distribución por áreas	219
Fig. 3.4 Errores tipográficos: distribución de tipos por áreas	221
Fig. 3.5 Número de palabras por búsqueda.....	222
Fig. 3.6 Tipos de búsquedas.....	223
Fig. 3.7 Número de palabras/tipo de búsquedas	223
Fig. 3.8. Número de palabras: distribución por áreas	228
Fig. 3.9 Tipos de búsqueda: distribución por áreas.....	229
Fig. 3.10 Tipos de coincidencia: distribución total (con correcciones)	235
Fig. 3.11 Número de palabras/Tipo de coincidencia.....	240
Fig. 3.12 Tipo de búsqueda/Tipo de coincidencia	241
Fig. 3.13 Estructura/Tipo de coincidencia	243
Fig. 3.14 Áreas temáticas: tipo de coincidencia.....	244
Fig. 5.1 Técnicas de recuperación de información: propuesta de secuenciación.....	285
Fig. 5.2. Ejemplo visualización de índices: coincidencia exacta y parcial	287
Fig. 5.3. Ejemplo visualización de índices: coincidencia parcial	289
Fig. 5.4. Ejemplo visualización de índices: coincidencia algún término en el índice.....	290
Fig. A.1 Errores tipográficos: distribución por áreas.....	368
Fig. A.2 Coincidencia exacta: tipos	374
Fig. A.3 Coincidencia parcial: tipos.....	374
Fig. A.4 Coincidencia exacta: número de palabras.....	375
Fig. A.5 Coincidencia exacta: tipo de búsqueda.....	375
Fig. A.6 Coincidencia exacta: estructura	376

Índice de tablas

Tabla 2.1 Transacciones campo materias INNOPAC/Millennium.....	187
Tabla 2.2 Distribución de las búsquedas por áreas de conocimiento.....	189
Tabla 2.3 Tipos de coincidencia/Técnicas de recuperación/Indices	210
Tabla 3.1 Características de las búsquedas/Errores tipográficos	218
Tabla 3.2 Búsquedas no válidas: tipos por áreas.....	220
Tabla 3.3 Tipo de estructura de los enunciados de búsqueda	226
Tabla 3.4 Estructura: distribución por áreas	231
Tabla 3.5 Coincidencia de las expresiones con los términos del índice	233
Tabla 3.6 Porcentaje corrección de errores sobre coincidencia	234
Tabla 3.7 Coincidencia parcial Índice/Enunciados.....	236
Tabla 3.8 Coincidencia palabras: palabras del enunciado en el mismo encabezamiento	236
Tabla 3.9 Coincidencia exacta y parcial/características: palabras y tipo.....	239
Tabla 3.10 Coincidencia exacta y parcial/características: estructura	239
Tabla 3.11 Coincidencia exacta por áreas. Prueba de Chi-cuadrado	245
Tabla 3.12 Coincidencia parcial por áreas. Prueba de Chi-cuadrado.....	246
Tabla 4.1. Errores tipográficos: datos comparados.....	250
Tabla 4.2. Campos de búsqueda: datos comparados.....	253
Tabla 4.3 Coincidencia enunciado/índice: datos comparados	255
Tabla 4.4 Tipo de coincidencia enunciado/índice: datos comparados	256
Tabla 4.5 Normalización por categorías de coincidencia: datos comparados.....	257
Tabla 4.6 Tipo de coincidencia enunciado/índice por áreas	267
Tabla A.1 Errores tipográficos: distribución por características.....	367
Tabla A.2 Búsquedas no válidas por áreas y tipos.....	367
Tabla A.3 Búsquedas no válidas: diferencias entre áreas. Prueba Chi-cuadrado	368
Tabla A.4 Errores ortográficos por áreas	368

Tabla A.5 Errores tipográficos: distribución por áreas	369
Tabla A.6 Errores tipográficos: diferencias entre áreas. Prueba Chi-cuadrado	370
Tabla A.7 Número de palabras: porcentaje total.....	370
Tabla A.8 Tipos de búsqueda.....	370
Tabla A.9 Tipos de búsqueda: número de palabras	371
Tabla A.10 Número de palabras: distribución por áreas.....	371
Tabla A.11 Número de palabras: diferencias entre áreas. ANOVA	371
Tabla A.12 Tipos de búsqueda: distribución por áreas	372
Tabla A.13 Tipos de búsqueda: diferencias entre áreas. Prueba Chi-cuadrado	372
Tabla A.14 Estructura: diferencias entre áreas. Prueba Chi-cuadrado.....	372
Tabla A.15 Coincidencia palabras en el índice.....	373
Tabla A.16 Coincidencia palabras en un encabezamiento.....	373
Tabla A.17 Porcentajes coincidencia	373
Tabla A.18 N° de palabras/ coincidencia	375
Tabla A.19 Tipo de búsqueda/Coincidencia exacta y parcial	376
Tabla A.20 Estructura/Tipo de coincidencia.....	377
Tabla A.21 Ciencias sociales: coincidencia de las expresiones con el índice.....	378
Tabla A.22 Exactas-Naturales: coincidencia de las expresiones con el índice.....	379
Tabla A.23 Humanidades: coincidencia de las expresiones con el índice	380
Tabla A.24 Ciencias médicas: coincidencia de las expresiones con el índice	381
Tabla A.25 Opac-Web: coincidencia de las expresiones con el índice.....	382
Tabla A.26 Tipo de coincidencia por áreas. Porcentajes	383
Tabla A.27 Coincidencia exacta por áreas. Prueba de Chi cuadrado.....	383
Tabla A.28 Coincidencia parcial por áreas. Prueba de Chi cuadrado	383
Tabla A.29 Áreas: características/coincidencia exacta. Prueba Chi-cuadrado	383
Tabla A.30 Áreas: características/coincidencia parcial. Prueba Chi-cuadrado.....	384

A quienes resisten sin olvidar los sueños.

“... Pero viajar no es un empeño en busca de lo imaginado, no es la persecución de algo que uno quiere ver, cerrando los ojos a todo lo demás. No es un deporte hecho para los que están seguros de lo que son, qué quieren y adónde van. Una sola pregunta puede justificar un gran viaje y el viaje está hecho para aquellos que no saben muy bien hacia dónde se dirigen ni conocen con exactitud lo que buscan. Está hecho para los que intuyen que encontrar no es lo importante y que cumplir un sueño puede ser, sobre todo, darse de bruces con la aventura...”

Javier Reverte.

“...Hacemos cosas, pero contarlos es difícil porque falta lo más importante, la ansiedad y la expectativa de estar haciendo las cosas, las sorpresas tanto más importantes que los resultados, los fracasos... Contar lo que hacemos es apenas una manera de rellenar los huecos inevitables...”

Julio Cortázar.

Agradecimientos.

A José Antonio, director de esta tesis, por todo lo que he aprendido a su lado a lo largo de estos años.

A Montse y a Asun, de la Biblioteca de la Univeridad de Vigo, sin cuya calurosa acogida este trabajo nunca habría echado a andar. A Angeles y a Inmaculada, de la Biblioteca de la Universidad Complutense de Madrid, que me facilitaron la información que necesitaba dedicándome parte de su tiempo.

A Alfonso, a Emilio y a José Luis, por sus buenos consejos, su tiempo y su generosa aportación a este trabajo.

A los compañeros de Facultad, por animarme y por recordarme que esto se acababa. Especialmente a Luis, quien por estar más cerca ha necesitado más paciencia para aguantar mis cambios de humor. También a Cris, a quien le ha tocado escuchar y leer este trabajo antes y más veces que a nadie.

A mi familia y a mis amigos, por estar ahí, comprendiendo la ausencia sin explicarse muchas veces la razón y confiando siempre en que algún día terminaría.

Introducción

Los estudios sobre la utilización de los catálogos en línea se iniciaron en la década de los ochenta en el ámbito anglosajón. Sus resultados demostraron el interés de los usuarios por las búsquedas por materias, a pesar de los problemas – conceptuales y mecánicos – que encontraban para plantearlas. Para intentar solventar estas dificultades, en todos ellos se proponían mejoras que facilitaran este tipo de acceso.

En la actualidad, los catálogos en línea de acceso público son habituales en las bibliotecas públicas o universitarias. Además, las redes de información han hecho posible que sean accesibles para cualquier usuario final de cualquier lugar del mundo. De esta manera, la “población de usuarios” de los OPAC’s es cada vez más numerosa y heterogénea, tanto en las habilidades con que se enfrentan a los sistemas como en lo que demandan de ellos. Un desarrollo lógico de esta evolución de la tecnología debería conducir a que cualquiera que necesite información sea capaz de acceder a ella por sus propios medios.

Sin embargo, a pesar de los avances tecnológicos y de casi dos décadas de investigación, muchos de los problemas que se constataron en la década de los

ochenta siguen teniendo vigencia en la actualidad: búsquedas que no recuperan registros, sobrecarga de información y escasa coincidencia del vocabulario utilizado por los usuarios con el empleado en los índices del sistema.

Estos problemas persisten porque los OPAC's han sido diseñados para facilitar búsquedas cuando el punto de acceso (autor, título, etc.) es conocido. Sin embargo, los estudios sobre el comportamiento de búsqueda de los usuarios han demostrado que el principal problema de la recuperación de información, y el que más incidencia tiene en la búsqueda por materias, es la dificultad que estos tienen para concretar su necesidad de información. Esta dificultad afecta tanto al proceso mediante el cual expresan su carencia de información en un planteamiento de búsqueda, como a la transformación de esa expresión en una estrategia adecuada a la estructura, vocabulario y sintaxis de sistema.

Para solventar estos inconvenientes los OPAC's deberían incorporar técnicas que mejoren los resultados de las búsquedas por equiparación (analíticas). Pero además deberían combinarlas con herramientas que faciliten las búsquedas exploratorias (browsing). En estas últimas el esfuerzo cognitivo se reduce porque el usuario sólo debe reconocer la información que se presenta en pantalla. Además, el planteamiento inicial (habitualmente muy genérico) se va modificando gracias a la información recibida durante la interacción con el sistema.

En España, a pesar de la popularización de la utilización de los catálogos en línea durante la última década, apenas se han realizado estudios empíricos sobre las características de las búsquedas por materias ni sobre el comportamiento los usuarios cuando las realizan. La comprobación de este vacío, junto con el interés por la indización y la recuperación de información (fruto de varios años de docencia en la materia), son el punto de partida de este trabajo.

En él se han analizado los enunciados de búsquedas por materias empleados por los usuarios del OPAC de la Universidad Complutense de Madrid. El objetivo era conocer sus características, las posibles diferencias entre distintas áreas de conocimiento y su grado de coincidencia con los índices temáticos. El análisis de estos factores debería permitir proponer medios para mejorar tanto el proceso como el resultado de las búsquedas.

Para empezar, se presenta un estado de la cuestión que sirve para contextualizar la investigación empírica. Una primera parte se dedica al proceso de búsqueda de información atendiendo al comportamiento de los usuarios, los modelos teóricos de los tipos de búsqueda y los problemas específicos de la recuperación por materias. Después, se describen las peculiaridades de los OPAC's para este tipo de búsqueda desde una triple perspectiva: la información temática que contienen, los índices que se crean con esta información y la forma de visualizar los resultados. A continuación, se recogen las principales propuestas de mejora relacionadas con el contenido de los registros del catálogo, las técnicas de recuperación, organización

de la información mediante estructuras de conocimiento y la interfaz. Por último, se presentan las principales corrientes epistemológicas sobre la materia y se analizan las metodologías empleadas en el estudio del comportamiento de búsqueda de los usuarios.

Por lo que respecta a las publicaciones empleadas en esta revisión bibliográfica, se ha recurrido fundamentalmente a tres fuentes. En primer lugar, a las revistas de mayor factor impacto en nuestra área de conocimiento (*Journal of Documentation, Journal of the American Society of Information Science, Information Processing & Management, Library & Information Science Research...*). También se han utilizado con frecuencia aquellas de mayor especialización en temas relacionados con el proceso técnico (*Cataloging & Classification Quarterly, Library Resources & Technical Services...*). Finalmente, se ha utilizado la información recogida en las actas de congresos y jornadas organizados por entidades o asociaciones relacionadas con la materia (IFLA, OCLC, ASIS, ISKO...), disponibles generalmente en la Web. Estos criterios explican el sesgo anglosajón de la bibliografía – justificado también porque es el ámbito geográfico en el que más se ha investigado sobre la materia – aunque siempre se ha procurado buscar publicaciones editadas en nuestro país.

Una vez perfilado el contexto teórico, se pasa a presentar la investigación empírica. Los objetivos aparecen acompañados de una serie de hipótesis de trabajo planteadas a partir de la lectura de investigaciones relacionadas, que han servido

como punto de partida para este estudio. La descripción de la metodología se ha considerado una parte fundamental del trabajo, porque la selección de los materiales y de los métodos de análisis determina el desarrollo de cualquier investigación. Por esta razón se detalla minuciosamente tanto el proceso de recogida y selección de información como las variables y métodos de análisis que se han empleado.

En el apartado de resultados se ha procurado exponer solamente los datos obtenidos del análisis. Sin embargo, se han añadido algunas explicaciones y ejemplos para aclarar algunos aspectos puntuales. Los comentarios de los resultados se exponen en el apartado de la “discusión”. Además se comparan estos con los de otros estudios similares, lo que ha permitido comprobar las hipótesis de partida y relacionar esta investigación con otras similares.

Las conclusiones ofrecen una síntesis de los resultados obtenidos, planteándose qué consecuencias pueden tener tanto las características como los niveles de coincidencia terminológica y sintáctica de los enunciados para la implementación de técnicas de búsqueda analíticas y exploratorias. Se han seleccionado aquellas propuestas de mejora que parecen más adecuadas en función de los resultados obtenidos. Para exponerlas se presentan ejemplos de la visualización de resultados de búsqueda que responderían a diferentes grados de similitud enunciado/índice y a la utilización de una combinación de técnicas de recuperación.

Esperamos que el trabajo realizado resulte innovador y que sus carencias sirvan como punto de partida para futuras investigaciones¹.

¹ Este trabajo ha sido elaborado en el marco del proyecto de investigación “El acceso por materias en los catálogos en línea: terminología y comportamiento de los usuarios” financiado por la Consejería de Educación y Cultura de la Junta de Castilla y León y el Fondo Social Europeo (Orden de 27 de enero de 2003, B.O.C. y L. De 19 de febrero de 2003).

**1. La búsqueda por materias en los catálogos
en línea: estado de la cuestión.**

La búsqueda por materias probablemente sea uno de los problemas más veces planteados, y menos resuelto, en la bibliografía sobre los catálogos en línea. En este capítulo se presenta un estado de la cuestión sobre el tema desde una triple perspectiva: la búsqueda de información y los problemas que encuentra el usuario durante este proceso; el contenido informativo del catálogo, su normalización, organización y estructura; y los medios propuestos para mejorar tanto la interacción usuario-sistema como los resultados. Por último, se ha dedicado un apartado a analizar la metodología que se ha utilizado en las investigaciones sobre búsqueda de información en los OPAC's.

1.1 La búsqueda de información en el OPAC.

La aparición de los catálogos de acceso público en línea fue el resultado del primer gran esfuerzo de la comunidad bibliotecaria para facilitar el acceso a la información a los usuarios finales gracias al uso de la tecnología. A mediados de la década de los

sesenta Swanson ya planteaba la posibilidad de que las bibliotecas automatizaran sus catálogos y que los usuarios emplearan directamente los terminales para buscar información bibliográfica (Su 1994). La idea de Swanson dejó de ser un proyecto a finales de los setenta, cuando las bibliotecas americanas empezaron a implementar sus OPAC's. Pocos años después la ALA los definía como:

“... un sistema informático que se utiliza para buscar en una base de datos bibliográfica desde diferentes terminales. Su diseño permite que los usuarios de la biblioteca puedan buscar y recuperar registros bibliográficos sin la ayuda de ningún intermediario...” (ALA Glossary of Library and Information Science 1983).

Como señalan Yee y Layne (1998), la automatización de los catálogos para facilitar su consulta no modificó sustancialmente los objetivos definidos por Cutter en 1876: (1) identificar o localizar un documento concreto y (2) conocer qué obras hay de un autor, las ediciones o versiones de una obra o qué publicaciones tratan sobre una materia. De hecho, estas funciones se mantienen en la actualidad y han servido a la IFLA como base para seleccionar los elementos básicos que deben contener los registros bibliográficos si quieren satisfacer las necesidades informativas de los usuarios (*Functional requirements for bibliographic records*, 1998). Con este propósito se realizó una redefinición de objetivos del catálogo, que según Svenonius (2000) moderniza y racionaliza planteamientos anteriores, en la que se considera que los usuarios lo utilizan para:

- ✓ Encontrar materiales que se correspondan con los criterios de búsqueda planteados: puede tratarse de un solo ítem o de un conjunto que comparte algún atributo común.
- ✓ Identificar una entidad, comprobando que se atiene a los criterios especificados o diferenciándola de otras entidades con algunas características similares.
- ✓ Seleccionar una entidad adecuada para su necesidad.
- ✓ Conseguir o acceder a los materiales seleccionados.

Resulta evidente que, pese a la persistencia de los objetivos, los OPAC's han cambiado considerablemente al ritmo de las innovaciones tecnológicas. Esta transformación ha supuesto el paso desde los sistemas de ámbito local, que emulaban los catálogos de fichas, hasta los actuales OPAC basados en la arquitectura cliente/servidor y adaptados al entorno gráfico e hipertextual de la Web (Beheshti 1997; Gorman 2000). Los cambios que se han producido durante estos años han afectado tanto a la información a la que pueden acceder los usuarios como a las técnicas de recuperación o a la interfaz.

Como ya pronosticaba Potter a finales de los ochenta (Potter 1989), se ha producido una ampliación del ámbito de los catálogos en línea relacionada con el incremento de la información a la que facilitan acceso. Este incremento ha sido tanto cuantitativo como

cualitativo ya que no sólo ha aumentado el número de registros de las bases de datos sino que se ha diversificado la naturaleza de los materiales incluidos en la colección.

La migración de los catálogos al entorno Web ha supuesto una aceleración de este proceso. Por una parte, porque simplifica la incorporación de recursos electrónicos a los que se puede proporcionar acceso directo, utilizando la etiqueta 856 del formato MARC para recoger la URL de cualquier tipo de documento digital (texto, imagen, sonido, etc.). Por otro lado, porque facilita la interconexión de sistemas, lo que ha permitido la unificación de colecciones locales y la conexión a otros catálogos o bases de datos (Delsey 2000; Marcos Mora 1999).

Actualmente, el desarrollo del estándar ANSI/NISO Z39.50 permite plantearse la creación de catálogos colectivos virtuales al normalizar el proceso para que distintos sistemas informáticos intercambien información, independientemente del software que utilicen (Evans 2001; Rosa Piñero, Senso Ruíz y Eito Brun 1998). De esta forma, desde la interfaz del ordenador cliente se puede acceder a recursos distribuidos en varios servidores porque es posible interrogar simultáneamente la base de datos de diferentes catálogos (López de Sosoaga 1998, 1999).

En cuanto a las herramientas de búsqueda, los actuales OPAC-Web fusionan las características de las generaciones anteriores (Hildreth 1984, 1985, 1989a). Al igual que

la primera, permiten realizar búsquedas por un campo completo (autor, título, materia, etc.), y como la segunda, incorporan opciones de búsqueda tradicionalmente empleadas en bases de datos en línea: palabras clave, operadores booleanos y truncamientos. La mayoría incluyen distintos tipos de delimitadores para que los usuarios especifiquen ciertas características de la búsqueda como tipo de documento, fecha de publicación, lengua, etc. (Hildreth 1995a; Moscoso 1998; Rodríguez Bravo y Alvite Díez 2001).

En el entorno Web se han añadido además enlaces hipertextuales que ofrecen la posibilidad de acceder de forma sencilla a diferentes documentos relacionados por una característica común, como puede ser el título, el autor, la materia o el código de clasificación. No obstante, se trata de “falsos” enlaces ya que las características del formato MARC impiden que se desarrolle una verdadera estructura de nodos relacionados. Lo que realmente está haciendo el sistema es ejecutar una búsqueda por equiparación en el índice correspondiente al campo que se selecciona (Ortiz Repiso y Moscoso 1999).

Una de las transformaciones más evidentes es la que afecta a la interfaz. Su diseño ha evolucionado, en primer lugar, desde los sistemas basados en órdenes a los sistemas de menús y formularios que son más apropiada para los usuarios inexpertos porque tiene la ventaja de proporcionarles un número limitado de opciones para elegir dentro de su

proceso de toma de decisiones¹ (García Marco 1995; Espelt 1998). Durante la década de los noventa los OPAC's se han adaptado al entorno Web implementando interfaces gráficas de usuario (GUI's), botones de selección de función, cuadros de diálogo y enlaces hipertextuales (Delsey 2000; Savage y Belkin 1999).

También se han incorporado mejoras para visualizar y manejar los resultados. La mayoría de los OPAC's actuales ofrecen la posibilidad de visualizar los registros bibliográficos en varios formatos (ISBD, etiquetas, MARC). Así mismo, es frecuente que incluyan opciones que permiten hacer reservas para el préstamo y seleccionar los registros visualizados para imprimirlos o enviarlos por correo electrónico.

A pesar de estos avances, no son pocos los autores que consideran que muchas veces se han realizado simples "mejoras cosméticas" (Hildreth 1995). Quienes se manifiestan en este sentido consideran que, aunque se han desarrollado en un entorno tecnológicamente avanzado, los actuales OPAC-Web no pueden considerarse realmente innovadores porque su estructura fundamental sigue reflejando los principios básicos de la tradición bibliotecaria. En cambio, para diseñarlos se ha tenido pocas veces en cuenta las características y necesidades del público al que se dirigen. Como se verá a continuación, los catálogos en línea siguen siendo difíciles de utilizar porque

¹ No obstante, los usuarios siguen necesitando entender las etiquetas de los campos y conocer los valores permitidos en cada uno de ellos (Estivill Rius, Gascón y Urbano 1995).

mantienen un paradigma de búsqueda obsoleto que no incorpora suficiente información sobre el comportamiento de búsqueda del usuario final.

1.1.1 Comportamiento de búsqueda de los usuarios.

Los resultados de las investigaciones han demostrado que los usuarios de los OPAC's realizan las búsquedas de forma intuitiva, sin ninguna o con muy poca planificación, emplean estrategias simples y no utilizan las herramientas avanzadas del sistema (Dillon 2000; Dimitroff 1992). A la hora de evaluar los resultados no son muy exigentes, quedan satisfechos con un número relativamente pequeño de registros relevantes y consideran que los sistemas actuales son más fáciles de utilizar que sus predecesores (Akeroyd 1990, Altuna Esteibar 1992, Hildreth 1995).

Pese a esta impresión general, que puede parecer favorable, estos mismos resultados evidencian que los usuarios tropiezan con bastantes problemas para realizar una búsqueda eficaz (Frías 1997, O'Brien 1994, Yee 1991). Encuentran dificultades para entender el funcionamiento del sistema con el que interactúan, para elegir los términos de búsqueda adecuados (Lancaster 1991; MacGarry y Svenonius 1991; Slone 2000) y para utilizar la lógica booleana cuando tienen que combinar conceptos (Ensor 1992; Hildreth 1997). En caso de que no recuperen registros no saben si se debe a que el

enunciado contenía errores ortográficos o tipográficos, a que no han utilizado el término correcto o a que no existe esa información en la base de datos. Normalmente abandonan la búsqueda si el sistema no les proporciona algún tipo de ayuda. Cuando recuperan demasiada información son incapaces de reducir eficazmente el conjunto de registros recuperados y visualizan un reducido número de ítems de ese conjunto (Blair 1990; Wiberley et al. 1995). También tienen problemas para entender las descripciones bibliográficas que se les presentan, de manera que valoran con dificultad la relevancia de los documentos que visualizan.

Por tanto, los OPAC's actuales pueden ser más "divertidos" de utilizar que los catálogos de las primeras generaciones, pero todavía presentan importantes limitaciones. Para Borgman (1996) esta situación es consecuencia de que han sido planteados como un reflejo de la estructura de los catálogos sobre fichas por lo que repiten su mismo esquema conceptual. Sin embargo, no se han adaptado a formas de buscar más dinámicas, aleatorias o fortuitas, que son más frecuentes entre los usuarios inexpertos. Como el modelo apenas ha cambiado en estos años, para interactuar con un catálogo en la era de la biblioteca digital un usuario sigue necesitando manejar las mismas habilidades y conocimientos que poco después de su aparición (Borgman 1986a, 2000):

- a) conocimiento conceptual para transformar la necesidad de información en una estrategia de búsqueda adecuada gracias a un correcto modelo mental del sistema.
- b) conocimiento semántico y sintáctico para poder implementar ese planteamiento de búsqueda (campos en los que puede buscar, tipos de búsqueda, operadores, etc.)
- c) habilidades técnicas básicas en el manejo del ordenador.

Actualmente, muchos usuarios tienen las habilidades técnicas necesarias para utilizar, con mayor o menor pericia, un ordenador. Este número se reduce considerablemente cuando se trata de los conocimientos conceptuales, semánticos y sintácticos necesarios para realizar una búsqueda relativamente satisfactoria. La mayor parte se aproximan al OPAC con unas expectativas poco realistas y, sobre todo, con modelos mentales inexactos del sistema. Este modelo conceptual es el que sirve como metaestructura al usuario en el proceso de transformación de su necesidad de información en un enunciado de búsqueda. Evidentemente, si el modelo no es correcto será una fuente de constantes problemas (Borgman 1986a; Crawford, Tom y Powles 1993).

Los nuevos OPAC's han reducido la complejidad sintáctica y, en alguna medida, las ambigüedades semánticas de la búsqueda. Sin embargo, no contribuyen a transformar

las necesidades de información de los usuarios en planteamientos de búsqueda adecuados porque ni mejoran su modelo mental, explicitando con claridad qué contienen y cómo está organizado, ni incorporan suficiente conocimiento sobre el proceso de búsqueda.

En opinión de Hildreth (1995), más que sugestivas interfaces gráficas, los usuarios necesitan ayuda para entender el proceso de búsqueda de información como una actividad consistente en diferentes niveles de interacción, complejos e interrelacionados, entre una variedad de tipos de datos y funciones. Este proceso se ha plasmado en diversos modelos teóricos que intentan sintetizar el conjunto de factores que intervienen: un *usuario* con una *necesidad de información*, que surge en un *contexto* concreto, cuya resolución pasa por la consulta de un *sistema* del que se espera obtener un *resultado* que resuelva el problema inicial (Marchionini 1997, Wilson 1997, 1999). Todos ellos coinciden en considerarlo un proceso inductivo e interactivo en el que todos los elementos están sujetos a modificaciones como consecuencia de esa interacción. Precisamente, la complejidad de una situación de búsqueda se debe a la indefinición, incertidumbre y variabilidad inherente a los diferentes niveles de intercambio de información que se producen (Kuhlthau 1991, 1993, 1998).

Uno de los modelos que mejor integra los diferentes elementos de este proceso es el propuesto por Saracevic (1996). En su “modelo estratificado de interacción” los

factores que intervienen – entorno, situación, conocimiento, búsqueda, interfaz, recursos informáticos e información – se articulan en una estructura simplificada en tres niveles: de superficie, cognitivo y de situación.

- ✓ En el nivel de superficie, un usuario interactúa con un sistema mediante una interfaz utilizando órdenes o preguntas que representan, de alguna manera, un problema de información. Al mismo nivel, el sistema responde tanto con meta-información, como con documentos (texto, imágenes, sonidos, etc.) o con preguntas que el propio sistema elabora para conseguir explicitar las características del problema del usuario.

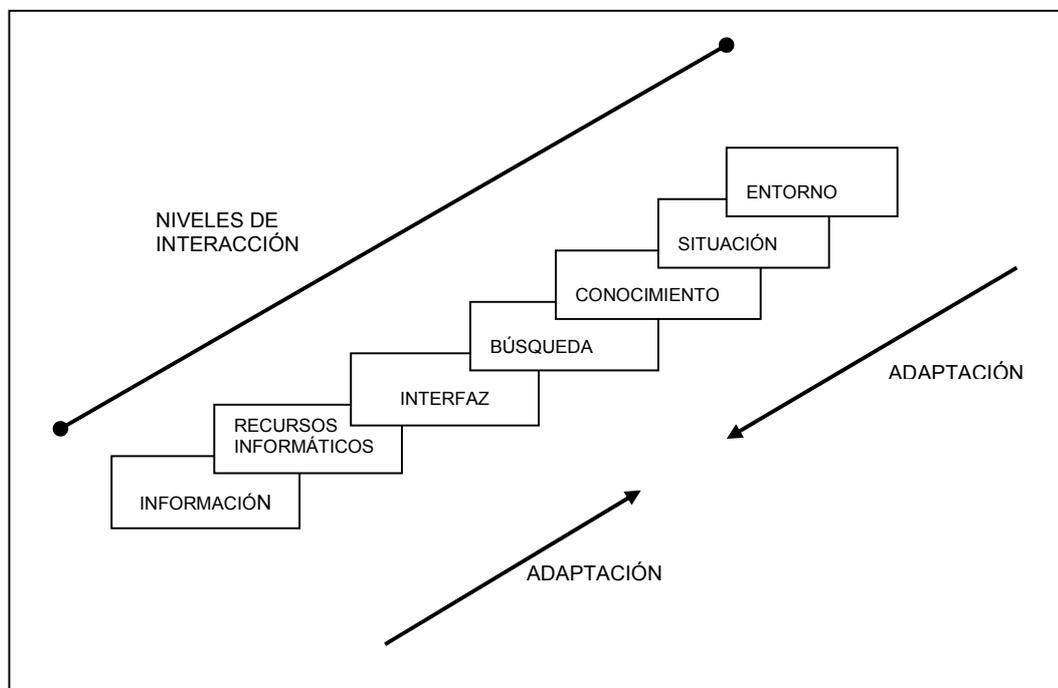


Fig 1.1 Modelo estratificado de interacción (Adaptado de Saracevic 1996a)

- ✓ En el nivel cognitivo, el usuario interactúa con la información que le ofrece el sistema. La operación que realiza es una evaluación de la utilidad de los documentos, en la que influyen tanto la información que se le facilita como su propio conocimiento del problema inicial.

- ✓ En el nivel de situación, el usuario interactúa con la situación o con el problema que ha producido la necesidad de información. El resultado de la búsqueda puede tener, o no, como consecuencia la resolución parcial o total del problema.

Los actuales sistemas se centran en la interacción en el nivel de superficie, en el planteamiento de la búsqueda y la visualización de resultados. Incorporan pocas herramientas diseñadas específicamente para que el usuario entienda cómo funciona el sistema, aclare su necesidad de información, interprete la información que visualiza o reconduzca con facilidad la búsqueda.

La investigación sobre el comportamiento de los usuarios durante las diferentes fases del proceso de búsqueda es la que puede ayudar a mejorar el diseño de los sistemas. Por ejemplo, para diseñar sistemas más flexibles interesa conocer si las características individuales de los usuarios o la claridad con que esté definido el problema influyen en

el tipo de estrategia de búsqueda (Borgman 1989, Slone 2000). También interesa saber si durante la interacción cambia el objetivo de la búsqueda, cómo lo hace, o si la resolución del mismo problema necesita varias sesiones de consulta (Hert 1996; Spink 1996; Spink et al. 1998). Finalmente, para elegir los formatos de visualización más adecuados habría que conocer qué factores influyen en los juicios de relevancia que realizan los usuarios sobre los documentos recuperados (Barry 1994, Park 1994, Peis y Fernández Molina 1994, Saracevic 1996a).

1.1.2 Métodos de búsqueda.

Los métodos de búsqueda de información se han agrupado tradicionalmente en torno a dos paradigmas: las **búsquedas analíticas**² y las **búsquedas exploratorias**³. Ambas técnicas o paradigmas de recuperación responden, con frecuencia, a distintas necesidades de información: buscar un documento específico sobre el que se tiene alguna información concreta o bien conocer qué recursos existen sobre un tema concreto.

² Se ha utilizado la expresión *búsqueda analítica* para traducir el término *query*, empleado en la bibliografía en inglés para referirse a este tipo de búsquedas.

³ En este trabajo el término inglés *browsing* se ha traducido como *búsqueda exploratoria*. Se ha adoptado esta expresión porque no existe un término claramente equivalente en castellano y la palabra exploración era la que se repetía con más frecuencia en los campos semánticos de los términos con los que se ha traducido *browsing* en la bibliografía en castellano: ojeo, espiguelo, navegación, etc. Además,

Para Hildreth (1991) la raíz de los problemas de los OPAC's reside en que su diseño y funcionamiento predomina el modelo de búsquedas analíticas, que refleja un paradigma conceptual parcial e inadecuado de la actividad de recuperación de información. Para este modelo el usuario debe representar su necesidad de información mediante un enunciado de búsqueda que se equipara con la representación de los documentos en los índices. El resultado que se obtiene del sistema es un conjunto único de resultados presentado como una lista de documentos potencialmente relevantes. Formular estrategias de búsqueda de este tipo requiere gran cantidad de esfuerzo cognitivo porque:

- ✓El sistema no puede responder antes de que el usuario le plantee su búsqueda, por lo que este planteamiento juega un papel muy importante en el proceso.

- ✓El usuario, en principio, no puede ver y seleccionar físicamente una fuente de información, sino que tiene que utilizar algún tipo de representación de la misma para recuperarla.

- ✓Para representar la información que necesita, el usuario tiene que formular un enunciado de búsqueda cuya eficacia depende de que utilice los términos y la sintaxis aceptados por el sistema (vocabulario, operadores, campos, etc.).

consideramos que representa de forma adecuada el contrapunto conceptual de la *búsqueda analítica (query)*

Este tipo de formulaciones puede ser operativo cuando el punto de acceso (autor, título o encabezamiento de materia) es conocido. Sin embargo, no son funcionales cuando se trata de necesidades más ambiguas o difíciles de definir. En estos casos, es prácticamente imposible que un usuario “inexperto” formule una estrategia de búsqueda que satisfaga su necesidad de información.

De hecho, las investigaciones sobre experiencias de usuarios reales han demostrado que el sistema de búsquedas analíticas es cuestionable, porque a menudo no tienen suficientes recursos para expresar su necesidad de información. Esta evidencia indica que es necesario implementar herramientas que posibiliten que un planteamiento inicial, por muy genérico que sea, se vaya modificando durante la interacción con el sistema. Durante este proceso, el usuario no sólo va conociendo las características y estructura del catálogo, sino que también concreta su necesidad de información.

Las **búsquedas exploratorias** representan un modelo de búsqueda alternativo en el que, de una manera dinámica, el usuario sólo tiene que reconocer y seleccionar la información que le presenta el sistema. Esto significa una reducción del esfuerzo cognitivo que realiza, porque se trata de un proceso mucho más simple que conceptualizar o recuperar la información de la memoria (Chang y Rice 1993, Fernández Molina y Peis 1997, Marchionini 1997).

A las características de este modelo responden métodos de búsqueda indirectos como por ejemplo el “*pearl growing*” (Borgman, Moghdam y Corbett 1984), que arranca de uno o varios documentos relevantes conocidos para, a partir de ahí recuperar material relacionado con ese conjunto inicial. Otro ejemplo es el sistema “*berrypicking*” propuesto por Bates (1989) en el que, a lo largo de varias visualizaciones, el usuario va seleccionando los materiales que le presenta el sistema y que están relacionados por compartir alguna característica común.

Considerando los diferentes comportamientos observados en los usuarios, Marchionini (1997) distingue cuatro tipos de estrategias en las búsquedas exploratorias:

- ✓ Ojeo-reconocimiento: es fundamentalmente una actividad de reconocimiento en la que un objeto claramente representado para el usuario se compara con un conjunto de objetos precisamente definidos. (Por ejemplo, reconocer un documento conocido en una lista).

- ✓ Observación: el usuario asume que se encuentra en un entorno en el que es muy posible que encuentre la información que busca y se encuentra en una actitud de predisposición a reaccionar ante cualquier estímulo.

- ✓ Estrategia de navegación. Equilibra la influencia del usuario y del entorno: el entorno facilita la exploración proporcionando posibles rutas y el usuario dispone de cierto control para elegir la ruta que va a seguir.

- ✓ Estrategia de rastreo. Es una estrategia muy similar al reconocimiento, excepto en que tolera entornos poco estructurados. Por ejemplo, mientras se lee un texto relacionado con una materia, una estrategia de rastreo “descubre” conceptos relacionados con otra área de interés. Es una estrategia que depende de cómo hace el usuario las asociaciones de conceptos y cómo se representan éstas en el espacio de información.

Se han establecido diferentes tipos de búsqueda exploratoria en función de las estrategias utilizadas y de la mayor o menor definición del objetivo de búsqueda. Sintetizando las tipologías tradicionales propuestas por Herner (1970) y por Cove y Walsh (1988) se pueden distinguir tres tipos:

- ✓ Búsquedas exploratorias dirigidas o específicas. Una actividad estructurada y dirigida en la cual el objetivo es conocido pero los criterios de búsqueda son imprecisos.

- ✓ Búsquedas exploratorias semidirigidas o de propósito general. Tienen un objetivo menos definido y se procede de manera menos sistemática. Por ejemplo, la consulta en una base de datos sobre un tema de interés utilizando algún término muy genérico y examinando los registros recuperados para ver qué se encuentra algún documento.

- ✓ Búsquedas exploratorias no dirigidas o de hallazgo fortuito. Se trata de una actividad puramente aleatoria, desestructurada y no dirigida que no tiene un objetivo consciente.

En cualquiera de estos casos, el centro de atención no es la elección del mejor término para buscar, sino conseguir que el usuario adquiriera la costumbre de explorar varios descriptores hasta encontrar el más adecuado para sus necesidades (Bates 1986). Por tanto, la preocupación por conseguir que el usuario encuentre el término exacto se sustituye por el diseño de herramientas que le muestren un abanico de términos y le ayuden a elegir dentro de esta variedad.

Por otra parte, la visualización un registro o registros recuperados no se considera el final de la búsqueda. En realidad, pueden ser el punto de partida si, por ejemplo, se utiliza la información de un documento conocido o seleccionado como relevante para recuperar otros relacionados. También puede ser un buen pivote para reconducirla, en

caso de que se localice algún tipo de información que haga que el usuario cambie de idea sobre lo que prefiere o, simplemente, porque cambie la manera de expresar su necesidad.

La eficacia de las búsquedas exploratorias depende de que la información que contiene el sistema esté organizada de manera consistente en función de diferentes criterios o características (temas, autores, títulos...). Evidentemente, también es necesario que se muestren al usuario las agrupaciones y las relaciones que se producen como resultado de la aplicación de estos criterios. Como consecuencia de estos planteamientos se han ido incorporando a los OPAC's algunas opciones que facilitan la visualización de los índices, de manera que el usuario pueda hacerse una idea de la información que contiene el sistema, seleccionar términos de búsqueda o ir del término que ha utilizado al admitido. También se han añadido enlaces hipertextuales que vinculan los registros que tienen el mismo autor, pertenecen a la misma serie o tratan sobre la misma materia.

Sin embargo, la implementación de las innovaciones necesarias para fomentar las búsquedas exploratorias encuentra barreras relacionadas con varios elementos del sistema: la estructura y contenido del registro bibliográfico, las técnicas de recuperación y los métodos para poner de manifiesto las relaciones entre los elementos informativos (Ortiz Repiso y Moscoso 1999). Estos obstáculos deberían superarse si se

quieren alcanzar algunos objetivos de los catálogos de tercera generación que, según Hildreth (1995) o Norgard (Norgard et al. 1993), hace ya años deberían:

- Incorporar sistemas como los árboles de búsqueda que mejoran los resultados de la búsquedas analíticas.
- Guiar al usuario desde los términos que utiliza hacia las entradas normalizadas del sistema.
- Facilitar la navegación entre registros que comparten alguna característica en común.
- Informar al usuario en cada momento de manera sensitiva y adaptada a cada situación.
- Porporcionar la opción de “encontrar más como éste”.
- Mostrar documentos conforme a un ranking determinado u ofrecer otro mecanismo para convertir cualquier conjunto recuperado en subconjuntos pequeños y ordenados.
- Incluir alguna herramienta bien desarrollada para el análisis de los conjuntos de documentos recuperados.

1.1.3 Problemas de la búsqueda por materias.

La mayor parte de las dificultades identificadas en la recuperación de información en catálogos en línea están relacionadas específicamente con las búsquedas temáticas. Y es que si consideramos la búsqueda de información como un proceso de resolución de problemas, las búsquedas por materias son problemas más difíciles de resolver porque están mal definidos. Como consecuencia, requieren mayor esfuerzo cognitivo del usuario para desarrollar las operaciones que llegarán a resolverlos (Borgman 2000).

Las dificultades se inician en el momento en que el usuario tiene que utilizar sus propios recursos para definir una carencia de información, un estado anómalo de conocimiento (Belkin, Oddy y Brooks 1982). Además, debe ser capaz de expresarla en un lenguaje concreto, utilizando términos aceptados por un sistema que le resulta poco familiar. La realización de búsquedas por materias requiere poner en funcionamiento diferentes tipos de conocimiento: del área temática, de las características de los campos que se van a utilizar para buscar, del vocabulario que se ha utilizado para la indización y de las técnicas de búsqueda (operadores, delimitadores, etc.).

Estos inconvenientes no han impedido que constituyan una importante actividad en los catálogos en línea. Las primeras investigaciones sobre el uso de los OPAC's

demonstraron que las búsquedas por materias eran más demandadas en estos que en los catálogos manuales (Cochrane y Markey 1983; Hancock-Beaulieu 1987, 1990; Lipetz y Paulson 1987; Markey 1984; Matthews et al. 1983). Sin embargo, este interés inicial pareció decaer ante las dificultades que encontraban los usuarios y la cantidad de búsquedas que no recuperaban ningún registro (Kaske 1988; Larson 1991b). Para superar esta situación se buscaron alternativas como la recuperación por palabras clave en el registro, que significó una reducción del uso del campo materias. A pesar de este descenso, los resultados de estudios recientes reflejan porcentajes de búsqueda en este campo que oscilan entre el 30 y el 57% (Poo y Khoo 1997; Yee y Layne 1998).

No obstante, ni los avances tecnológicos, ni la popularización de las redes de información, ni más de dos décadas de investigación han conseguido evitar muchos de los problemas de la búsqueda temática, que ya se constataron en la década de los ochenta y que siguen teniendo vigencia en la actualidad (Hunter 1991; Husain y O'Brien 1992; Martínez y Pichinini 1996; Olson y Boll 2001; Poo y Khoo 1997).

El primer grupo de problemas se encuentra en el acceso al sistema. En este paso el usuario debe conseguir formalizar y expresar de alguna manera su necesidad de información. Para hacerlo adecuadamente, necesitaría tener un modelo mental claro de ese sistema con el que interactúa: contenido, campos de búsqueda, información de cada campo, etc. (Borgman 1986a; Dimitroff 1992). Es difícil que el usuario pueda hacerse

esta idea, si lo que encuentra cuando se acerca a un sistema de búsqueda es un cuadro de diálogo que, la mayoría de las veces, sólo le permite elegir un campo de búsqueda en el que debe teclear algo.

Varios trabajos coinciden en señalar una conducta similar respecto a las búsquedas por materias en estas circunstancias: los usuarios tratan de expresar los temas de búsqueda de forma escueta, utilizando enunciados muy breves (una o dos palabras). También se observa que comienzan utilizando términos mucho más generales que los que reflejarían su necesidad de información, para posteriormente ir refinando o especificando los resultados (Drabenstott 1996a; Hancock-Beaulieu 1987; Jones, Cunningham y McNab 1998; Pulis y Ludy 1988; Walker y Jones 1987).

Otro aspecto que plantea dificultades es el vocabulario: cómo encontrar los términos adecuados para enunciar la búsqueda y para modificarla en caso necesario. Se han estudiado las características de los términos y expresiones empleados por los usuarios (Bates 1993; Bilal 1998), su coincidencia con los encabezamientos de materia del catálogo (Bates 1977; Carlyle 1989; Drabenstott 1996a; Drabenstott y Vizine-Goetz 1994; Markey 1984), la incidencia de los errores ortográficos o tipográficos (Drabenstott y Weller 1996; Ferl y Millsap 1996; Zink 1991), los problemas para encontrar términos alternativos que permitan replantear la búsqueda (Connell 1995; Efthimiadis 1995; Fidel et al. 1999; Lancaster et al. 1991) y las secuencias de

movimientos realizados en la modificación (Ferl y Millsap 1996; MacGarry y Svenonius 1991; Spink y Saracevic 1997).

Los resultados de estas investigaciones han demostrado que los usuarios utilizan enunciados de búsqueda bastante simples, de los cuales entre un 4% y un 6% contienen errores ortográficos o tipográficos. Además, menos de un 40% de ellos coinciden exactamente con los encabezamientos del índice de materias. En este último caso, existen muchos problemas para utilizar otras opciones de búsqueda o encontrar términos alternativos y, por lo general, se abandona el proceso.

Evidentemente, el vocabulario es el aspecto más difícil de la búsqueda: conceptos, lugares, ideas y objetos son denominados de una manera diferente por quien crea los documentos, quien los indiza y quien los busca (Collantes 1995). Los expertos utilizan tesauros u otras herramientas para identificar sinónimos, términos más genéricos que mejoren la tasa de llamada y términos específicos para incrementar la precisión (Fidel 1991). Sin embargo, el usuario “inexperto” de un OPAC se va a encontrar, probablemente sin saberlo, en la siguiente situación:

- ✓ Si realiza la búsqueda en un campo de materias “normalizado”, es difícil que conozca suficientemente los códigos de clasificación o la terminología y la sintaxis de los encabezamientos para realizar búsquedas con éxito. También es

difícil que se pueda aprovechar de la estructura conceptual que existe en estas herramientas porque, por un lado, la desconoce y no sabe utilizarla y, por otro, pocas veces está implementada en los sistemas de forma clara y manejable.

- ✓ Si utiliza un campo de búsqueda en el que se emplea el lenguaje natural (título, notas, resumen, etc.) es difícil que conozca y utilice las técnicas que emplearía un profesional: sinónimos para un mismo concepto, especificaciones para diferentes significados de un mismo término, truncamientos para evitar las diferencias entre formas flexionadas, variantes ortográficas, etc.

Por otra parte, también existen dificultades para manejar las técnicas de recuperación implementadas en los OPAC: operadores booleanos, operadores de proximidad, delimitadores, etc. A partir de la segunda generación se incorporaron estas capacidades de búsqueda, habituales en las bases de datos bibliográficas, que resultan eficaces en manos de profesionales pero no de los usuarios finales. Por ejemplo, la utilización de operadores booleanos no ha supuesto una mejora sustancial, porque los usuarios no entienden cómo emplearlos y confunden su significado lógico con el de estos términos en lenguaje natural⁴, lo que provoca combinaciones erróneas. De hecho, como señalaba Borgman (2000), los usuarios de los catálogos en línea son reacios a la utilización de

⁴ Por ejemplo, el Y lógico booleano significa una intersección de los conjuntos de documentos que contienen los términos: gatos Y perros, los documentos recuperados tratarán de las dos especies. En el lenguaje natural la conjunción Y puede tener sentido tanto de intersección como de yuxtaposición de los conceptos, aunque es más habitual este último: gatos y perros, tanto una especie como la otra.

herramientas complejas, cuya relación con la tarea que ejecutan es abstracta y poco evidente, por lo que generalmente no utilizan las opciones de búsqueda más avanzadas de los catálogos.

Los resultados obtenidos se mueven entre las búsquedas que no recuperan registros – más de un tercio – y las que provocan sobrecarga de información (Larson 1991, 1991b). Como ya se ha señalado, cuando no se recuperan registros o se recuperan pocos los usuarios no saben cómo incrementar los resultados y cuando se han recuperado demasiados no saben cómo reducirlo.

Pese a lo complicado del panorama, los usuarios no parecen desalentarse y siguen utilizando los OPAC's para hacer búsquedas por temáticas, aunque cabe preguntarse ¿hasta cuándo? y ¿cómo cambia esta situación cuando ese mismo usuario se desenvuelve en entornos informáticos mucho más amigables? (Frías 1997). Como hemos visto en el apartado anterior, muchas de estas dificultades se deben a la persistencia de un paradigma de recuperación obsoleto, basado en búsquedas analíticas y que en muchos casos no ha tenido en cuenta al usuario. Las investigaciones han demostrado que las necesidades de información que no están claramente definidas, como suelen ser las temáticas, se resuelven mejor mediante búsquedas exploratorias. En este tipo de búsqueda es fundamental la interacción entre el usuario y el sistema, entendida como un diálogo, una negociación durante la cual se va concretando la

necesidad de información, se va conociendo el sistema y se van seleccionando documentos relevantes.

Estos mismos trabajos demuestran que existen alternativas tanto para facilitar el proceso como para mejorar los resultados. Para conseguirlo, es necesario enriquecer las estructuras de datos, modernizar las técnicas de recuperación y favorecer la interacción sacando partido de las estructuras de conocimiento y de las opciones de visualización. Algunas de las soluciones son sólo experimentales, pero otras ya se han incorporado a los sistemas comerciales, aunque no se incluyen entre las opciones básicas y su implementación incrementa el coste del sistema, por lo que muchas veces queda supeditada a necesidades presupuestarias (Seguí i Palou y Vall 1999).

1.2 La información temática en los catálogos en línea.

Cuando un usuario utiliza el OPAC para hacer una búsqueda por materias puede emplear dos tipos de información (Markey 1984, Yee y Layne 1998, Peters y Kurth 1991):

- ✓Una información no normalizada, formada por el contenido de diferentes campos del registro bibliográfico que, aunque no han sido diseñados para ser puntos de acceso temático, contienen datos que denotan, en mayor o menor medida, el tema de los documentos.

- ✓La información específicamente temática y normalizada que se incluye en los campos para los códigos de clasificación y encabezamientos de materia.

A continuación se examinará con más detalle cada una de estas opciones, especialmente los campos que contienen información específica sobre materias. Dentro de ellos se prestará más atención a los encabezamientos de materia por dos razones: 1) es el campo más utilizado por los usuarios para realizar búsquedas temáticas, y 2) es el apartado que tiene más relación con la investigación empírica de este trabajo.

1.2.1 Información temática no normalizada.

Aunque cualquier campo del registro bibliográfico es potencialmente útil para ayudar en la búsqueda temática, aquellos que tienen un contenido más rico son título y notas:

- ✓El título muchas veces proporciona terminología que no está presente en el campo materias y que actualiza o especifica el vocabulario que se puede encontrar en éste.

- ✓La variedad de notas que se contemplan en el formato MARC⁵ hacen que este campo sea, al menos potencialmente, una rica fuente de información: nota de tesis (502), de bibliografía (504), de contenido (505), de citas o referencias bibliográficas (510), de sumario (520), etc. Aunque en la práctica su uso es muy limitado, muchas veces no sólo aportan información alternativa o más específica, sino que se precisa el contenido temático del documento. Un ejemplo son las notas del campo 505 en las que es posible hacer una descripción analítica de las aportaciones individuales que se incluyen en un volumen bajo un título común, a las que se suele asignar una materia general.

```
01 BNE19950172459
005 199901070931
008 950602m19951998esp|///|/||| ||spa/j
019 $a B 9399-1995
020 $a 84-320-8370-4
040 $a M-BN $b spa $c M-BN
080 0 $a 94(460)
130 00 $a Historia de España
245 1 $a Historia de España $c dirigida por Antonio Domínguez Ortiz
260 0 $a Barcelona $b Planeta $c 1995-1998
300 $a 12 v. $b il. col. $c 28 cm
504 $a Bibliografía
505 0 $a 1. Desde la prehistoria hasta la conquista romana (siglo III a.C.) $a 2. La España romana y visigoda (siglos III a.C.-VII d.C.) $a 3. Al-Andalus: musulmanes y cristianos (siglos VIII-XIII) $a 4. De la crisis medieval al Renacimiento (siglos XIV-XV) $a 5. El siglo de Oro (siglo XVI) $a 6. La crisis del siglo XVII $a 7. El reformismo borbónico (1700-1789) $a 8. Descubrimiento, colonización y emancipación de América $a 9. La transición del antiguo régimen (1789-1874) $a 10. La restauración (1874-1902) $a 11. Alfonso XIII y la segunda república (1902-1939) $a 12. El régimen de Franco y la transición a la democracia (de 1939 a hoy). Índice general
651 8 $a España $x Historia
700 21 $a Domínguez Ortiz, Antonio $d 1909-2003
```

Fig. 1.2 Registro bibliográfico. Formato IBERMARC

Para explotar estos datos se utiliza la búsqueda por palabras clave dentro del registro. La desventaja de este tipo de búsqueda es que para que sea correcta deben neutralizarse algunos de los problemas que plantea el lenguaje natural: sinonimia, polisemia, homonimia, variantes flexionadas, etc. Otro inconveniente es que la información que contiene el registro bibliográfico en estos campos es muy limitada, por lo que, como veremos, algunas propuestas insisten en la necesidad de enriquecer el contenido de los registros para mejorar la búsqueda por materias.

⁵ En adelante, la información se referirá a las etiquetas que identifican los campos en el formato IBERMARC (2002). Sólo se indicarán aquellos casos en que la información se refiera a un formato

1.2.2 Encabezamientos de materia.

La información específica sobre el contenido temático de los documentos se recoge en los campos materias del registro bibliográfico. La estructura de este campo en el formato MARC (6XX) responde tanto al tipo de materia que se trate – persona, lugar, tema, etc. – como a las peculiaridades del vocabulario controlado que se utiliza para normalizar estos puntos de acceso: las listas de encabezamientos de materia.

Campos 6XX Encabezamientos de materia	Identificadores de subcampo específicos para 6XX
600 Nombre de persona como materia	\$a Encabezamiento
610 Nombre de entidad como materia	\$j Subdivisión de forma
611 Nombre de congreso como materia	\$x Subdivisión general de materia
630 Título uniforme	\$y Subdivisión cronológica
650 Materia	\$z Subdivisión geográfica
651 Nombre geográfico	
653 Términos de materia no normalizados	

Fig. 1.3 Campos IBERMARC para encabezamientos de materia

En los campos 600-651, el primer indicador denota características específicas de cada tipo de encabezamiento y el segundo el sistema empleado para normalizar el punto de acceso temático⁶. Respecto a los identificadores de subcampo, se refieren tanto a las

MARC diferente.

⁶ Una descripción exhaustiva del formato MARC queda fuera del alcance de este trabajo. Sin embargo, se ha considerado importante exponer las características esenciales de los campos de contenido temático – fundamentalmente los que recogen los encabezamientos de materia - por la relación que tienen con la parte experimental del mismo. La información que se presenta es siempre un breve extracto del

peculiaridades de los nombres de personas, entidades, congresos o títulos uniformes (estos son similares a los empleados en los campos 1XX correspondientes) como a las subdivisiones que habitualmente se utilizan para especificar los encabezamientos de materia.

600 0# \$aCarlos \$bIII \$c, Rey de España \$x- Pensamiento político
610 2# \$aComisiones Obreras \$b. Unión Regional de Asturias \$x-Historia
611 2# \$aExpo'92 \$c(Sevilla) \$x- Instalaciones
630 0# \$aCasablanca (Película cinematográfica)
650 ## \$aAutomóviles \$x- Diseño \$y- Italia
651 ## \$a Córdoba \$x- Historia \$y- 1931-1939 \$j- Congresos y Asambleas

La normalización de los campos 600, 610, 611 y 630 se basa fundamentalmente en las reglas establecidas en los códigos catalográficos de los diferentes países, mientras que la que aparece en 650 y 651 está normalizada en la lista de encabezamientos empleada en cada centro.

La idea de describir el contenido de los documentos utilizando términos normalizados se remonta a mediados del siglo XIX. Su primer defensor fue Crestadoro (1856) quien señaló la posibilidad de utilizarlos en los catálogos impresos como complemento a las palabras del título. El primero en utilizar los encabezamientos de materia fue Abott en

contenido del Formato IBERMARC para registros bibliográficos (2002), Formato IBERMARC para registros de autoridad (1999) y MARC 21 cuando no existe equivalente español.

el catálogo del Harvard College en 1861 (Hanson y Daily 1970) y las primeras normas para su redacción fueron las que estableció Cutter (1876).

Una vez que se decidió utilizar términos normalizados para representar las materias de los documentos, fue necesario crear listas con ellos para facilitar la consistencia del vocabulario empleado por distintos catalogadores. La necesidad de que esta consistencia se extendiera a diferentes centros para favorecer el intercambio de información justifica la aparición de los primeros repertorios de materia. Este es el origen de listas como la *Library of Congress Subject Headings*⁷ o el *Répertoire de vedettes-matière de la Universidad de Laval* que constituyen repertorios de referencia, no sólo para sus países de origen sino para las lenguas inglesa y francesa.

En 1994 la sección de Clasificación e Indización de la IFLA inició un proyecto de investigación cuyo objetivo era elaborar un estado de la cuestión sobre las listas de encabezamientos empleadas en las bibliografías nacionales o en los catálogos de las bibliotecas nacionales (Heiner-Freiling 2000). El análisis de los datos de 88 países demostró un claro predominio de la utilización de la LCSH (en inglés o traducidas). El desarrollo de importantes catálogos de autoridad en las bibliotecas nacionales al margen de este repertorio es una característica típica de los 90. Sistemas tan importantes como RAMEAU (*Répertoire d'autorité matière encyclopedique et*

⁷ En adelante LCSH

alphabetique unifié), RSWK (*Regeln für den Schlagwortkatalog*) y SIPORBASE (*Sistema de Indexação em Português*) se han desarrollado durante esta década como un producto de la creciente importancia de la indización en los servicios centrales nacionales basada en progreso de las tecnologías de la información y de la utilización de OPAC's.

La situación en nuestro país es bien diferente. Aunque se han realizado importantes traducciones de la LCSH a la lengua española – por ejemplo *Bilindex* – no se ha elaborado un repertorio propio equiparable a los que acabamos de citar. Ante esta perspectiva, las bibliotecas han adecuado los encabezamientos de materia a sus necesidades, generándose numerosos sistemas *ad-hoc*. En la actualidad, las bibliotecas públicas del Estado español utilizan mayoritariamente la *Lista de encabezamientos de materia para bibliotecas públicas* (1995) y las *Autoridades de la Biblioteca Nacional* (1996-). El panorama en las bibliotecas universitarias es mucho más heterogéneo: aunque el repertorio más utilizado -26,33%- es la lista *de Encabezamientos de materia de la Universidad de Sevilla* (1992-), el 39,54% de los centros elaboran sus propios encabezamientos traduciéndolos de otras listas⁸ o redactándolos a partir de varias fuentes (Caro Castro y Travieso Rodríguez 2003a).

⁸ predominantemente de LCSH (77,78% de las traducciones) aunque también de RVM LAVAL (22,11% de las traducciones).

Las consecuencias más evidentes de esta variedad son la falta de consistencia en la representación de las materias y la inexistencia de una estructura semántica coherente y eficaz. Estos dos factores dificultan las búsquedas tanto en el OPAC del propio centro, como en los catálogos colectivos distribuidos que permite crear el protocolo Z39.50 (Caro Castro y Travieso Rodríguez 2003; Gil Leiva 2001, Jiménez Rodríguez 1998).

La falta de consistencia agrava uno de los grandes problemas del acceso temático en los catálogos en línea: la no coincidencia entre los términos empleados por los usuarios y los encabezamientos de materia. El motivo de esta disparidad se atribuye habitualmente a la limitación del vocabulario, y a una normalización y una sintaxis que muchas veces resultan extrañas al uso “natural” del lenguaje (Foskett 1997; Martínez y Pichinini 1998; Rowley 1994)

En los encabezamientos de materia, como en cualquier otro vocabulario controlado, se realiza una normalización de los términos para limitar su significado. La regulación afecta tanto a la morfología como a las relaciones semánticas y a la sintaxis. El objetivo es conseguir un lenguaje en el que cada término represente un solo concepto (evitando la polisemia, homonimia y homografía) y cada concepto esté representado por un solo término (controlando la sinonimia). El nivel más elemental de normalización es el que determina la categoría gramatical de los descriptores y regula la forma que van a tener

los términos en cuanto a número o género. En el extremo opuesto, el nivel más complejo es el que se refiere a la sintaxis.

1.2.2.1 Sintaxis.

Las reglas sintácticas en el lenguaje natural se refieren al orden de las categorías de palabras (nombres, verbos, adjetivos, etc.) y a los nexos que pueden unirlos para formar sintagmas (nominales, verbales, adverbiales) y frases. En los vocabularios controlados sucede algo similar. Un encabezamiento de materia puede estar formado por un solo elemento o por varios elementos encadenados pertenecientes también a diferentes categorías (Encabezamiento – Subencabezamientos). Cada uno de estos elementos representa un concepto que puede expresarse mediante una o varias palabras. La sintaxis en este caso también determina tanto la composición de cada uno de los elementos como el orden y los nexos que permiten crear una cadena con elementos de diferentes categorías (Svenonius 2000).

Como se acaba de indicar, cada encabezamiento o subencabezamiento puede estar formado por una sola palabra, preferentemente un sustantivo, o por un sintagma nominal que puede tener la forma “nombre + adjetivo” o “nombre + sintagma preposicional”. No obstante, el uso de los sintagmas nominales está limitado en los

lenguajes controlados con el objetivo de que sólo expresen un concepto y esto muchas veces significa diferencias respecto a lo que están acostumbrados a utilizar los usuarios (*UNE 50-106-90 Directrices para la elaboración de tesauros monolingües; Normativa para la redacción de encabezamientos de materia 1991*).

Por ejemplo, los adjetivos se utilizan en principio como en el lenguaje natural para diferenciar clases o tipos de un concepto más genérico (*Bibliotecas infantiles*). Sin embargo, su uso no siempre está permitido: cuando se trata de gentilicios u otras formas que significan origen o procedencia sólo se pueden utilizar con sustantivos que se refieran a manifestaciones culturales o artísticas, a grupos de personas fuera de su país de origen o para especificar lenguas y literaturas. Esto último significa que, por ejemplo, se puede emplear la forma *Arte francés* pero no *Elecciones europeas*. Evidentemente, esta limitación resulta extraña a la forma en que normalmente se expresan los usuarios.

En cuanto a la forma “sustantivo + complemento preposicional” sucede algo parecido. Se trata de una forma admitida siempre que sirva para especificar un concepto o cuando se trata de una denominación comúnmente admitida y no cuando su uso suponga la unión de dos conceptos en un mismo sintagma. Por ejemplo, se puede utilizar *Hospitales para niños* o *Comités de empresa*, pero no se puede emplear

Migración de las aves (Migración + Aves) o *Catalogación de manuscritos* (Catalogación + Manuscritos).

En los casos en que para representar la materia de un documento es necesario combinar varios conceptos se crea una cadena en la que se combinan varios elementos simples (encabezamiento y subencabezamientos). Esta normalización da lugar a la inversión del “orden directo” y/o a la eliminación de los nexos de algunos sintagmas en el lenguaje natural

Expresión	Encabezamiento
Migración de las aves	Aves - Migración
Catalogación de manuscritos	Manuscritos - Catalogación
Conservación de los alimentos	Alimentos - Conservación
Tesis sobre medio ambiente	Medio ambiente - Tesis
Diseño de automóviles	Automóviles - Diseño
Historia de Madrid	Madrid - Historia

Fig. 1.4 Encabezamientos de materia. Sintaxis. Combinación de conceptos

Las normas sintácticas también concretan cuándo se pueden utilizar los subencabezamientos y en qué orden. La mayor parte de los trabajos realizados en este sentido intentan establecer un orden lógico que justifique la sucesión de los términos. Las reglas al respecto se remontan a los principios pragmáticos establecidos por Cutter (1876):

- ✓ El encabezamiento principal es la entrada más específica sobre la que trata el documento. Esto impide la sistematización dentro del encabezamiento que debe ser alfabético y directo (*Gatos*) y no sistemático e indirecto (*Animales domésticos – Gatos*).

- ✓ El orden será Lugar-Materia para aquellos encabezamientos relacionados con la historia, gobierno o comercio de una zona geográfica (*Córdoba – Historia*).

- ✓ Se utilizará la forma “nombre +adjetivo” en el encabezamiento en el caso de la literatura y el arte (*Literatura gallega*).

Sin embargo, la mayoría de los trabajos tratan de superar este pragmatismo buscando unos principios teóricos que sustenten normas sintácticas de carácter más o menos permanente. Siguiendo el modelo propuesto por Kaiser (1911), Ranganathan (1933, 1951) Farradane (1950, 1952, 1963), Coates (1960) y Austin (1977), la sintaxis se define con relación a una serie de clases de términos que son homogéneos: entidad, proceso, propiedades, lugar, tiempo, etc. (Fugmann 1993; San Segundo Manuel 1999 ; Svenonius 2000)

En el caso de los encabezamientos de materia, la definición de las categorías de términos se hace de acuerdo a criterios tanto semánticos como funcionales, lo que

complica la sintaxis (*Principles underlying subject headings* 1999). El resultado es un primer conjunto de términos de materia que pueden funcionar como encabezamientos, cuando denotan el tema principal del documento, o como subdivisiones de materia cuando se refieren a procesos, propiedades, características, etc. del tema representado por el encabezamiento. A estas dos categorías se suman las de lugar, tiempo y forma.

Los términos pertenecientes a estos grupos se combinan en cadenas de acuerdo a un orden que nunca ha sido demasiado estable. Las construcciones sintácticas más habituales son (Chan 1995; *Lista de encabezamientos de materia para bibliotecas públicas* 1995; Svenonius 2000a):

Encabezamiento principal – Lugar – Materia – Tiempo – Forma

Encabezamiento principal – Materia – Lugar – Tiempo - Forma

Encabezamiento principal de nombre geográfico – Materia – Tiempo – Forma

Pese a este predominio, la variedad de combinaciones era tan amplia que en la década de los ochenta la Library of Congress empezó a plantearse la necesidad de regularizar esta situación en la LCSH (*Subject cataloging manual: Subject headings* 1985). El resultado fue un conjunto de reglas sintácticas tan complejas que no resulta extraño que en 1991 se celebrara una conferencia con el objetivo de discutir la posibilidad de simplificarlas (Chan 1992; Conway 1992).

En primer lugar se abordó la cuestión de si los subencabezamientos eran necesarios en un entorno digital: ¿en aras de la simplificación no se deberían separar en campos diferentes? Finalmente se decidió mantener la estructura, porque además de permitir especificar la materia y contextualizar términos ambiguos, no debería ser un obstáculo para la recuperación en sistemas automatizados.

También se reconocieron los beneficios de una sintaxis regular. Los participantes propusieron que se estandarizara el orden de los subencabezamientos bajo los encabezamientos temáticos: *Encabezamiento – Materia – Lugar – Tiempo – Forma*. Sin embargo, se consideró que esta propuesta no era viable en algunos casos porque suponía modificar el significado de determinados encabezamientos. Se trataba especialmente de cadenas empleadas en las disciplinas de arte, literatura, historia y derecho: *Música – Brasil – Historia y crítica* puede significar algo completamente diferente a *Música – Historia y crítica – Brasil*.

Así mismo, se abordó el tema de si estos matices sintácticos eran perceptibles para los usuarios. Aunque se realizó un estudio para comprobarlo (Franz et al. 1994; Drabenstott, Simcox y Fenton 1999), los resultados no fueron concluyentes y finalmente se tomó la decisión de mantener la sintaxis que se utilizaba aunque el precio fuera la complejidad. La justificación fue que el precio de mantener la sintaxis, aunque sea compleja, no es tan alto cuando los sistemas automatizados permiten recuperar

cualquier término del encabezamiento, cualquiera que sea el lugar en el que se encuentre. A cambio se consigue mayor especificidad.

1.2.2.2 Estructura semántica.

En el ámbito semántico en los vocabularios controlados se establecen relaciones entre los términos que reproducen el entramado conceptual de uno o varios campos de conocimiento. Estas relaciones quedan reflejadas en un sistema de reenvíos, presente tanto en el lenguaje de indización como en el catálogo, que constituye la estructura sindética del sistema de indización (Davies y Rush 1979; *Principles underlying subject headings* 1999). Estas relaciones tienen dos objetivos:

- ✓ Eliminar la ambigüedad del lenguaje natural, evitando la dispersión “de elementos afines” (Lancaster 2002). Esto se consigue gracias al control de la sinonimia mediante las relaciones de equivalencia y a la diferenciación del significado de los términos polisémicos u homónimos utilizando calificadores.

Marionetas
V. Títeres
Títeres
Up. Marionetas

Diafragma
Diafragma (Fotografía)
Diafragma (Anticonceptivo)

✓ Crear una red con los descriptores que reproduce las vinculaciones semánticas entre ellos. En esta red están presentes dos tipos de relaciones:

- Jerárquicas: Ponen de manifiesto la relación entre dos descriptores de los que uno está subordinado a otro por carácter normativo. Pueden ser de dos tipos:

Género/Especie (Genéricas): *Fruta* → *Manzanas*

Todo/Parte (Partitivas): *Aparato locomotor* → *Esqueleto*

- Asociativas: Se definen por negación como aquellas relaciones que se establecen entre términos que se asocian, dentro de un campo de conocimiento, sin que la relación que los vincula sea ni jerárquica ni de equivalencia (UNE 50-106-90). Para complementar esta vaga definición se han fijado unos criterios que determinan las categorías de términos entre los que se puede establecer relaciones asociativas: una disciplina y su objeto de estudio (*Sismología* → *Terremotos*), un proceso y su resultado (*Catalogación* → *Catálogos*), un proceso y el objeto sobre el que recae (*Clasificación* → *Documentos*), productos y sus propiedades (*Venenos* → *Toxicidad*), etc.⁹.

Las listas de encabezamientos de materia incorporan estas relaciones, aunque su estructura semántica ha sido criticada por sus deficiencias: la estructura jerárquica es imperfecta y muchas veces está incompleta, las relaciones asociativas no responden a otro criterio que el puro azar y, además, el sistema de reenvíos no permite diferenciar la relación jerárquica de la asociativa (Coates 1960; Foskett 1997; Fugmann 1993; Svenonius 2000). Sistemas como la LCSH han intentado corregir estas insuficiencias racionalizando las relaciones mediante criterios claramente establecidos, similares a los empleados para la elaboración de tesauros (Chan 1995a; Drabenstott y Vizine-Goetz 1994; Olson y Boll 2001). También se ha mejorado el sistema de reenvíos: a partir de la undécima edición, las referencias “va” y “xx” se convierten en siglas que identifican los términos genéricos (BT), específicos (NT) y relacionados (RT).

Animales domésticos	Animales domésticos
x Animales de carga	UP Animales de carga
xx Construcciones ganaderas	TG Animales
Industria agopecuaria	TE Animales de compañía
Veterinaria	Aves de corral
Animales	Ganado
va Construcciones ganaderas	RT Construcciones ganaderas
Industria agopecuaria	Industria agopecuaria
Veterinaria	Veterinaria
Animales de compañía	
Aves de corral	
Ganado	

Fig. 1.5 Encabezamientos de materia. Estructura semántica

⁹ Se pueden encontrar ejemplos adicionales en: *UNE 50-106-90 (1990)*, Lancaster (2002), Foskett (1997).

1.2.2.3 Registros de autoridad de materia.

Para controlar la forma de los términos y rentabilizar las relaciones semánticas de cara a la recuperación de información en los OPAC's, es necesario gestionar adecuadamente el catálogo de autoridades de materia (Tartaglia 2003). Como señalan Olson y Boll (2001), los registros de este catálogo son la versión electrónica del vocabulario controlado que emplea el centro. Se redactan siguiendo las directrices elaboradas por el Comité de Clasificación e Indización de la División de Control bibliográfico de la IFLA (*Directrices para los registros de autoridad y referencia de materia* 1995¹⁰), y se codifican utilizando el Formato MARC de autoridades (Formato IBERMARC para registros de autoridad y referencia, 1999).

Los registros de autoridad pueden ser de dos clases:

- ✓ Registros de autoridad propiamente dichos, que contienen la información relativa a un encabezamiento autorizado y sus referencias asociadas.

¹⁰ Todavía no se han adaptado para materias ni traducido al castellano las *Guidelines for Authority Records and References* de 2001

- ✓ Registros de referencia, que sirven para dirigir al usuario desde un término no admitido a un término admitido o para conducirlo desde un término admitido a otros relacionados.

La información que contienen puede referirse también a dos tipos de términos: a) encabezamientos específicos de materia, término materia (150) y término geográfico (151), y b) términos que pueden utilizarse como puntos de acceso descriptivos o temáticos: 100 (personas), 110 (entidades), 111 (congresos), 130 (títulos uniformes). Además pueden redactarse registros de autoridad para las subdivisiones de materia (180) - generales o aplicables a una clase de términos¹¹ -, geográficas (181), cronológicas (182) o de forma (185).

Como puede apreciarse en el ejemplo, los registros del catálogo de autoridades contienen tanto las formas autorizadas de los encabezamientos como los reenvíos a sus equivalentes y a todas sus relaciones. Puede incluir además referencias comprensivas (Pueden usarse además...) y notas explicativas (Úsase para... Úsase también como...) que sirven como medio de orientación en el catálogo y aclaraciones de uso de las autoridades. Finalmente se pueden añadir notas que especifiquen la fuente utilizada, que describan la historia del encabezamiento o que aclaren su uso al indizador.

¹¹ Bajo nombres de personas, lugares, profesiones, etc.

150	##	\$aAnimales domésticos
360	##	\$iVéase además los nombres de los distintos animales domésticos: \$aPerros \$aGatos, etc.
450	##	\$aAnimales de carga
550	##	\$wg \$aAnimales
550	##	\$wh \$aAnimales de compañía
550	##	\$wh \$aAves de corral
550	##	\$wh \$aGanado
550	##	\$aConstrucciones ganaderas
550	##	\$aIndustria agropecuaria
550	##	\$aVeterinaria
670	##	\$aLCSH:\$b(Domestic animals)
150		Encabezamiento admitido (\$a)
360		Referencia compleja de “véase además” Frase explicativa (\$i) Encabezamientos a los que remite (\$a)
450		Referencia de véase Encabezamiento no admitido (\$a)
550		Referencia de véase además Términos relacionados (\$a) Puede especificarse si se trata de términos genéricos (\$wg) o específicos (\$wh)
670		Nota de fuente Fuente(\$a) Encabezamiento original(\$b)

Fig 1.6 Registro de autoridad y referencia comprensiva (IBERMARC)

Además de reproducir todos los datos relativos a un encabezamiento o subencabezamiento, los registros de autoridad se relacionan entre ellos y con los registros bibliográficos (Frías 1994; Jiménez Pelayo y García Blanco 2002). Esta información es muy valiosa desde el punto de vista del acceso temático porque puede dirigir al usuario desde sus propios términos a los empleados en el sistema o sugerirle

términos alternativos. Si se quiere conseguir que se utilice con eficacia para la recuperación debe trasladarse a la interfaz de consulta y hacerlo de manera que sea comprensible por el usuario. Desgraciadamente, esto sucede con poca frecuencia y sólo en los últimos años se está empezando a utilizar para la consulta una herramienta que tradicionalmente se consideraba de uso exclusivo de los bibliotecarios.

1.2.3 Códigos de clasificación.

Durante décadas, la comunidad bibliotecaria se ha mostrado más favorable al uso de vocabularios controlados para asignar términos de indización a los documentos que a utilizar los tradicionales esquemas de clasificación que agrupan los documentos en una estructura jerárquica de categorías temáticas. En esta situación, la función de las clasificaciones ha quedado limitada a la ordenación física de los documentos, y aún en este papel han tenido numerosos detractores (Dominguez Sanjurjo 1996).

Una muestra evidente de esta situación es la escasez de datos sobre los códigos de clasificación que se incluyen en el registro bibliográfico y la infrautilización que se hace de esta información para la recuperación. Aunque en el formato MARC21 aparecen doce campos (050/092) que pueden contener códigos de diferentes sistemas

de clasificación, con frecuencia sólo se utilizan uno o dos. Por ejemplo en el formato IBERMARC lo habitual es usar uno (080).

052 Código de clasificación geográfica
080 Número de la Clasificación Decimal Universal
081 Número de la Clasificación ICONCLASS
082 Número de la Clasificación Decimal de Dewey
084 Otras clasificaciones bibliográficas

Fig. 1.7 Formato MARC 21. Campos para códigos de clasificación

Además, la información que se incluye en este campo es muy escasa: se limita a la notación sin incluir ningún subcampo que permita explicar su significado (rúbrica). Es realmente difícil que un usuario pueda sacar partido de estos datos porque es poco posible que conozca el sistema de clasificación de forma suficiente como para introducir un código como criterio de búsqueda y no existen alternativas que permitan ir de las expresiones lingüísticas a los códigos. La única opción realmente funcional es utilizar el código de un registro relevante para recuperar registros relacionados, si es que existe algún tipo de enlace que permita hacerlo.

En el caso del formato IBERMARC el campo más utilizado es el 080 porque el sistema de clasificación que se utiliza mayoritariamente en nuestro país es la CDU. Sin embargo, esta aparente homogeneidad esconde muchas excepciones. Es cierto que los cincuenta años de uso obligatorio de este sistema (1939-1989) para elaborar el

catálogo sistemático de materias en la red de Bibliotecas Públicas del Estado lo han convertido en una referencia bastante estable en este ámbito (San Segundo Manuel 1996). Muy diferente es la situación en bibliotecas universitarias y de investigación donde los sistemas especializados, las adaptaciones y los esquemas creados *ad-hoc* conviven con sistemas tradicionales como la CDU y, cada vez más, la LCC (San Segundo Manuel, 1999a). Las consecuencias de esta heterogeneidad para la recuperación de información son similares a las que hemos señalado en el apartado anterior con los encabezamientos de materia.

Paradójicamente, esta infrutilización de los códigos de clasificación se produce a pesar de que desde la aparición de los primeros catálogos en línea se han realizado experimentos para demostrar el potencial de los principales sistemas de clasificación como instrumentos para mejorar el acceso temático. El primer proyecto importante fue el desarrollado Freeman y Atherton (1968). En él se demostró la capacidad de la *Clasificación Decimal Universal* (CDU) para mejorar la recuperación de información, aunque sus conclusiones no tuvieron una repercusión directa en la evolución ni en la utilización de este sistema (MacIlwaine 1997). En 1984 un equipo de la Universidad de Toronto, dirigido por Nancy J. Williamson, comenzó una investigación sobre las posibilidades de automatizar la *Library of Congress Classification* (LCC) para facilitar la recuperación en sistemas interactivos en línea (Williamson 1991). Las posibilidades de la *Dewey Decimal Classification* (DDC) se

investigaron en el *Dewey Decimal Classification Online Project*, patrocinado por OCLC y dirigido por Karen Markey (Markey y Demeyer 1986).

Las conclusiones de estos trabajos sirvieron para demostrar que la incorporación de un esquema de clasificación al catálogo en línea puede asegurar un acceso temático más eficaz, complementando las posibilidades del acceso alfabético. En síntesis, los argumentos a favor de su utilización han sido (Chan 1995; Iyer 1995a; Svenonius 1983):

- ✓ Facilitan las búsquedas exploratorias porque permiten la presentación en pantalla de los registros temáticamente afines, independientemente de que estos estén disponibles en el centro, estén prestados o pertenezcan a otras unidades de información. Esta organización virtual de la colección por categorías temáticas también permite que a partir del código asignado a un registro conocido se puedan identificar otros potencialmente relevantes porque tratan la misma materia.

- ✓ Se puede ampliar o limitar la búsqueda a conceptos más genéricos o específicos gracias a las posibilidades de navegación de la estructura jerárquica.

- ✓ Se pueden contextualizar las búsquedas dentro de un campo semántico. Esto resulta especialmente práctico cuando se combina con la búsquedas por palabras clave ya que elimina problemas de ambigüedad del lenguaje natural como la polisemia.

- ✓ Los códigos de clasificación pueden funcionar como un lenguaje puente para salvar barreras lingüísticas en bases de datos muy amplias con registros en diferentes lenguas.

1.2.3.1 Formato MARC para códigos de clasificación.

Las investigaciones que hemos reseñado no sólo pusieron de manifiesto la utilidad de las clasificaciones para la recuperación de información, sino la necesidad de formalizar y aclarar los datos de sus notaciones en los registros del catálogo. Una de las consecuencias de sus recomendaciones fue la creación del formato *USMARC para Códigos de Clasificación* que básicamente utilizaba la estructura del formato MARC de autoridades para codificar la información clasificatoria (*USMARC format for classification data* 1991). El formato se pensó en principio para automatizar la LCC – proceso que finalizó en 1995 (Guenter 1996) – aunque posteriormente se modificó

para adaptarlo a las notaciones de otros sistemas como de DDC o la CDU
(*Requirements for a format for classification data* 1996).

El resultado de estos esfuerzos es la reciente publicación del formato MARC 21 para datos de clasificación (2000-2002).

084	\$acdu\$International medium edition
153	\$a791.43 \$hBellas artes. Juegos. Espectáculos. Deportes \$hDiversiones. Espectáculos. Juegos. Deportes \$hEntretenimientos públicos. Espectáculos \$jCine. Películas
553 0	\$wl \$a688.74 \$hCiencias aplicadas. Medicina. Tecnología \$hIndustrias, oficios y comercio de artículos acabados o ensamblados \$hJuguetes. Artículos de diversión e ilusionismo \$hEquipo para teatros y circos. Aparatos de prestidigitación. Equipos para salas de variedades y parques de atracciones \$jEquipo para salas de cine. Accesorios de proyección.
553 0	\$wl \$a778.5 \$hBellas artes. Juegos. Espectáculos. Deportes \$hFotografía y procesos similares \$hAplicaciones y técnicas especiales de fotografía \$jCinematografía
553 0	\$wjh\$a791.43-92 \$hBellas artes. Juegos. Espectáculos. Deportes \$hDiversiones. Espectáculos. Juegos. Deportes \$hEntretenimientos públicos. Espectáculos \$jCine. Películas \$tPelículas documentales
750	\$aCine
750	\$aPelículas cinematográficas
001-008 Códigos de control	
084 Esquema de clasificación (\$a) y edición (\$c)	
15X Códigos de clasificación y rúbricas explicativas	
153 Código de clasificación (\$a), su rúbrica (\$j) y las de los niveles jerárquicos superiores (\$h)	
4XX Referencias a códigos modificados o eliminados.	
5XX Referencias a otros códigos de clasificación.	
553 Referencia equivalente a "véase además". Puede referirse tanto a códigos relacionados porque tratan la misma materia desde el punto de vista de otra disciplina (\$wl) como a códigos más específicos (\$wjh) o genéricos (\$wjg)	
70X-75X Términos de indización asociados	
Repite la estructura de los campos e identificadores utilizadas en los campos 6XX del registro bibliográfico: 700 Nombre de persona, 710 Nombre de entidad, 711 Nombre de congreso, 730 Título uniforme, 750 Materia, 751 Nombre geográfico, 753 Términos de indización no controlados.	

Fig. 1.8 Formato MARC 21 Para datos de clasificación

En este formato el registro de cada código contiene su propia rúbrica e información sobre la cadena jerárquica a la que pertenece. También incluye campos para notas de aplicación, referencias a números relacionados y obsoletos, y para establecer relaciones con encabezamientos de materia, descriptores y palabras clave. Como puede observarse en el ejemplo, el formato explota fundamentalmente la estructura jerárquica de los sistemas de clasificación, pero apenas ofrece posibilidades de aprovechar las facetas que incorporan sistemas como la *DDC* o la *CDU* en forma de auxiliares comunes y especiales (Riesthuis 1998).

A pesar de sus limitaciones, puede convertirse en una potente herramienta para facilitar búsquedas exploratorias siempre que la política de clasificación del centro permita sacar partido de todas sus posibilidades. Para Olson y Boll (2001) estas son:

- ✓ Permite el acceso alfabético por los términos que explican el significado de las notaciones y por todos aquellos pertenecientes a encabezamientos de materia, descriptores, etc. que se hayan vinculado al código. Por ejemplo los registros de los códigos de la edición electrónica de la *DDC*¹² incluyen enlaces a los encabezamientos de materia correspondientes de la *LCSH*.

¹² Aunque su automatización es anterior a la publicación del MARC 21 los formatos utilizados son compatibles y contienen información similar (Olson y Boll 2001).

- ✓ Permite visualizar términos más genéricos o más específicos. Muestra la posición relativa de una notación dentro de su jerarquía. Por ejemplo, la implementación de la LCC permite visualizar una jerarquía en una ventana mientras la notación en cuestión está en la ventana principal y permite el salto de un número a otro directamente.

- ✓ Permite conocer aspectos relacionados porque se pueden incluir códigos que traten sobre la misma materia en otra disciplina.

1.2.4 Índices y opciones de búsqueda.

Los diseñadores de catálogos en línea han elaborado tradicionalmente cuatro tipos de índices con la información recogida en los registros bibliográficos y de autoridad. Los cuatro tipos no son excluyentes, de manera que un mismo sistema puede ofrecer opciones de búsqueda sobre diferentes tipos de índices. Como señalan Yee y Layne (1998), a pesar de esta reiterada presencia no existe una terminología claramente establecida para designarlos. Para evitar las ambigüedades que se pudieran derivar de esta situación la definición de cada uno de ellos va acompañada de ejemplos.

Los primeros en utilizarse, mimetizando los catálogos manuales, fueron los “Índices bibliográficos” o índices de encabezamientos (autor, materia y título). Son “índices de frase”: cada entrada contiene un encabezamiento completo o la información de un subcampo del formato MARC. Cuando se busca en uno de estos índices, la cadena de palabras del enunciado de búsqueda tiene que coincidir exactamente con el orden de los términos del índice.

Encabezamiento	650 08 \$aMujeres\$x Situación social\$zAndalucía
Índice campo completo	Mujeres-Situación social- Andalucía
Índice subcampos	Andalucía Mujeres Situación social

Una variante de estos son los índices permutados en los que cada campo constituye también un elemento de indización, pero se permutan todos los subcampos que lo forman. De esta forma se crean tantas entradas de índice como combinaciones se puedan realizar con las subdivisiones¹³. Por lo tanto, cada subcampo constituye un punto de acceso que se mantiene “en contexto” junto a toda la cadena. La comparación entre enunciado de búsqueda e índice continúa siendo exacta y secuencial (de izquierda a derecha).

¹³ Otra opción de este tipo de índice, de uso poco frecuente, es la rotación de las palabras en lugar de los subcampos.

Encabezamiento	650 08 \$aMujeres\$x Situación social\$zAndalucía
Índice permutado	Andalucía-Mujeres-Situación social Andalucía-Situación social-Mujeres Mujeres-Andalucía-Situación social Mujeres-Situación social- Andalucía Situación social-Andalucía-Mujeres Situación social-Mujeres-Andalucía

Otra posibilidad es la indización mediante claves de búsqueda. En este caso, cada campo o subcampo se considera como una entrada de índice, manteniendo el orden de los elementos, pero sólo se almacena un segmento fijo de cada una de las palabras que lo integran. Por ejemplo, se puede establecer la secuencia: cuatro primeras letras de la primera palabra, tres de la segunda, dos de la tercera y la primera letra de la cuarta.

Encabezamiento	650 08 \$aMujeres\$x Situación social\$zAndalucía
Índice clave de búsqueda	Mujesitsoa

Los índices de “palabras clave” se incorporaron a los catálogos en línea para ampliar las posibilidades de búsqueda introduciendo opciones habituales en las bases de datos bibliográficas. En ellos se incluyen todas las palabras¹⁴ de los campos y subcampos que se han considerado puntos de acceso: encabezamientos descriptivos, título, notas, materias, etc. Se puede crear un solo índice conjunto o generar uno para cada campo. Es habitual que se eliminen las palabras que no tienen significado léxico utilizando un antídicionario o diccionario de palabras vacías o bien mediante algún algoritmo que

¹⁴ Entendiendo por palabra cada una de las secuencias de caracteres separadas por espacios.

tenga en cuenta el número de caracteres o la frecuencia de aparición del término. Básicamente, la búsqueda en este índice se realiza sobre cada uno de los términos y sus posibles combinaciones mediante operadores booleanos o de proximidad.

La creación de este tipo de índices con los términos de los encabezamientos de materia permite mitigar los efectos de la rigidez de la precoordinación de estos sistemas. De esta manera se reduce el riesgo de que la búsqueda no recupere registros, pero al precio de incrementar el volumen de información recuperada y el riesgo de ruido documental (Bermello Crespo 2001)

Título	245 10\$aSer mujer en Andalucía \$cAmparo García Arévalo
Encabezamiento	650 08 \$aMujeres\$x Situación social\$zAndalucía
Índice palabras clave	Amparo Mujeres Andalucía Ser Arévalo Situación García Social Mujer

1.2.5 Presentación de los resultados.

En cuanto a la presentación de los resultados de las búsquedas, aunque varía notablemente de unos sistemas a otros las “respuestas” más habituales son el índice alfabético del campo en que se ha realizado la búsqueda o una lista de los registros recuperados (normalmente con formato autor/título).

Muchos catálogos en línea responden a la búsqueda por materias presentando una lista alfabética de todos los encabezamientos que empiezan con el término o expresión del usuario. Otros presentan una lista de encabezamientos que contienen la palabra o palabras utilizadas por el usuario como respuesta a una búsqueda por palabras clave en el índice de materias. Algunos ofrecen una función de “explorar materias”¹⁵ en la que como respuesta a una búsqueda (haya obtenido resultados o no) se sitúa al usuario en el lugar que correspondería al enunciado en el índice alfabético.

MATERIA				
		integración social	Ver Catálogo Completo	Bu
Núm	Marcar	MATERIAS (1-12 de 45)	Año	Entradas Encontradas: 178
1	<input type="checkbox"/>	Integración social		
2	<input type="checkbox"/>	Integración social -- África	1993	
3	<input type="checkbox"/>	Integración social -- América Latina		
4	<input type="checkbox"/>	Integración social -- Argentina	1998	
5	<input type="checkbox"/>	Integración social -- Catálogos	1993	
6	<input type="checkbox"/>	Integración social -- Congresos		
7	<input type="checkbox"/>	Deficientes -- Integración social		
8	<input type="checkbox"/>	Deficientes -- Integración social -- España	1998	
9	<input type="checkbox"/>	Deficientes físicos -- Integración social		
10	<input type="checkbox"/>	Deficientes mentales -- Integración social		
11	<input type="checkbox"/>	Integración social -- Ensayos	1998	
12	<input type="checkbox"/>	Integración social -- España		

Fig. 1.9 Presentación de resultados. Índice alfabético permutado

¹⁵ *Subject browse.*

Como ya hemos señalado, otra posibilidad es que se presente directamente el listado de los registros cuyos encabezamientos cumplen las condiciones de la búsqueda. Generalmente se visualizan en formato abreviado (Título/Autor) y, a veces, en un orden poco inteligible para el usuario. Finalmente, existe la alternativa de que la presentación varíe en función del tipo o del campo de búsqueda elegido (Lawrence 1985; Wool 2000).



Fig. 1.10 Presentación de resultados. Registro abreviado (Búsqueda: Integración social)

Actualmente, muchos sistemas de gestión bibliotecaria han flexibilizado la definición de los índices y permiten que sean los propios centros quienes determinen qué tipos van a utilizar y los campos o subcampos que van en cada uno de ellos. La elección del tipo

de indización que se va a emplear es un factor muy importante porque tiene un profundo efecto en el funcionamiento general del sistema (Frakes y Baeza-Yates 1992; Larson 1991a; Moya Anegón 1995; Yee y Layne 1998).

Por ejemplo, ya se ha señalado que los sistemas que ofrecen acceso por palabras clave habitualmente permiten la búsqueda en varios campos simultáneamente. Cuando se realiza este tipo de búsqueda es muy difícil organizar los resultados en función del encabezamiento recuperado, porque frecuentemente la equiparación se produce en más de un campo. También es complicada la ordenación cuando se busca en un solo campo, porque los elementos que han servido como criterio de búsqueda pueden aparecer en cualquier orden y en cualquier lugar dentro del mismo. Esto dificulta el diseño de pantallas con presentaciones adecuadas para visualizar los resultados de este tipo de búsquedas.

Por otro lado, los sistemas que ofrecen índices de campos completos o subcampos parecen asumir que cada búsqueda va a recuperar sólo un encabezamiento y que no será necesario especificar por otros campos. Además, es necesario que el usuario conozca las primeras palabras del encabezamiento, que coincidan con todas las del enunciado y en el mismo orden. Para contrarrestar esta rigidez, la mayor parte de los sistemas realizan un truncamiento a la derecha cuando ejecutan estas búsquedas, de manera que se recuperan todos los encabezamientos “que empiezan por...” Esto

incrementa notablemente el número de encabezamientos recuperados pero reduce el silencio documental, que es el mayor riesgo de este tipo de búsqueda.

En cualquiera de los dos casos parece fundamental que se desarrollen sistemas eficaces para la visualización de los encabezamientos recuperados. Visualizaciones que faciliten y estimulen la navegación, la exploración, con la finalidad de que el usuario vaya eligiendo entre las opciones que el sistema le presenta después de haber equiparado su enunciado de búsqueda con los índices.

1.3 Mejoras en el diseño de los catálogos en línea.

Peters (1991) señalaba hace años dos opciones básicas para solucionar los problemas de la búsqueda en los OPAC's: educar a los usuarios para que saquen el máximo partido al sistema o cambiar el sistema para adaptarlo a los usuarios.

La primera opción implica el desarrollo de programas de formación de usuarios con el objeto de mejorar el modelo mental que tienen del sistema. Ya hemos señalado que esta idea que el usuario tiene del sistema y de cómo funciona es primordial para que pueda elaborar un enunciado correcto, entender los errores que comete durante el

proceso de búsqueda y buscar soluciones acordes. Si este modelo no se lo proporciona el diseñador del sistema o un instructor, el usuario elabora un modelo mental basado en su experiencia anterior y en sus interacciones con el sistema que suele ser incompleto y/o incorrecto (Borgman 1986, 1996).

Las investigaciones sobre la incidencia de diferentes modalidades de formación en el comportamiento de búsqueda parecen demostrar que los mejores resultados se obtienen con métodos de formación estructurados. Además deben ser programas que puedan llegar a todos los usuarios sin grandes esfuerzos para la biblioteca, que les permitan aprender a su ritmo y que pongan el énfasis en enseñar conceptos y estructuras más que procedimientos (Fernández Molina y Moya Anegón 1998; Ríos García 1991).

Aunque la alternativa de la formación de usuarios tiene ventajas evidentes (Mano González 1994), la opción de modificar el sistema parece haber despertado más interés como tema de investigación (Peters 1991). También es la que se va a desarrollar aquí porque es la más relacionada con la parte empírica de este estudio.

Numerosos autores han sugerido que los catálogos en línea deberían ser mejorados para realizar búsquedas por materias más efectivas (Estivill i Rius 1984, 1986; Fernández Molina y Moya Anegón 1998; Fernández Molina y Peis 1995; Hildreth 1995a; Long 2000; Martínez y Pichini 1996; Poo y Khoo 1997). Las mejoras que se han propuesto pueden dividirse en las siguientes categorías:

- Flexibilizar la estructura y enriquecer el contenido de los registros bibliográficos
- Mejorar las técnicas de recuperación.
- Explotar las estructuras de conocimiento incorporadas a los sistemas.
- Diseño de interfaces más fáciles de entender y utilizar.

1.3.1 Estructura de los registros.

Uno de los mayores inconvenientes que se plantean hoy en día para mejorar el rendimiento de los catálogos en línea es la rigidez de los registros bibliográficos en formato MARC.

La estructura actual de los registros bibliográficos se basa en las ISBD y en la reglas de catalogación. Se diseñó para describir y acceder a la información de un registro

completo y no para establecer relaciones lógicas entre los elementos informativos que lo forman o entre diferentes registros de una base de datos. El formato MARC reproduce estas características, de manera que se adapta bien a la comunicación de la información bibliográfica pero plantea problemas para la manipulación automatizada de los datos. Su estructura, lineal y rígida, no sólo es difícil de entender por los usuarios; además se adapta mal a la información organizada jerárquicamente e impide el desarrollo de una estructura de nodos y enlaces inherente al entorno Web (Gorman 1992; Frías y Ríos Hilario 2002; Svenonius 2000).

Las alternativas se orientan hacia la búsqueda de formatos más flexibles que faciliten la creación y presentación de las relaciones entre los distintos campos, así como la utilización de la estructura sindética del catálogo. Esto es especialmente importante para la búsqueda por materias, porque sin hacer uso de las relaciones semánticas que existen en los vocabularios controlados y en los sistemas de clasificación estos no dejarán de ser el obstáculo que hasta ahora se han considerado para el acceso temático. Además, al facilitar la creación de relaciones entre los campos se incrementarían las posibilidades de elaborar índices cruzados entre los campos materias, clasificación y título o notas, que ya han sido utilizados experimentalmente para mejorar el acceso por materias (Doszkocs 1983; Larson 1991).

La rigidez del formato MARC y de la estructura de la información no han hecho más que confirmarse cuando los catálogos han migrado a entornos Web y ha empezado a incorporar documentos electrónicos a sus fondos. De hecho, el papel del registro bibliográfico en un entorno digital todavía no está claro. Este entorno puede suponer cambios fundamentales porque permite tomar los datos directamente de los documentos, incorporarles información normalizada (metadatos) y, sobre todo, porque abre la posibilidad de una deconstrucción radical del registro bibliográfico a favor de los diferentes elementos que lo constituyen (Svenonius 2000).

Actualmente se plantea con frecuencia la posibilidad de codificar los registros bibliográficos en un formato diferente al MARC para conseguir hacerlos más abiertos e intercambiables y para facilitar la integración de recursos en la biblioteca digital. Frente a algunas visiones apocalípticas, autores como Tillet (1996; 2003) o García Camarero y García Melero (2001) piensan que en un futuro próximo el formato MARC continuará siendo el principal vehículo para comunicar información bibliográfica y registros de autoridad. Sin embargo, para obtener todas las ventajas de la evolución tecnológica es necesario un cambio, tanto en la filosofía como en la práctica de la catalogación, que permita manipular la información de un objeto bibliográfico único como múltiples objetos de información.

Estos factores han impulsado la tarea de diferentes instituciones y grupos de trabajo, que han abordado el problema dedicándose al diseño de reglas de catalogación y

formatos automatizados asociados que sean compatibles con normas de codificación como SGML/XML y con modelos de metadatos para la descripción de documentos electrónicos.

Por un lado, los esfuerzos se han centrado en la elaboración de modelos que permitan representar documentos de cualquier tipo y en cualquier soporte. El objetivo es crear un entramado genérico de datos elementales que sean compatibles entre las diferentes estructuras utilizadas para la descripción de documentos: desde las tradicionales reglas de catalogación hasta los actuales esquemas de metadatos (fundamentalmente Dublin Core¹⁶ y Text Encoding Initiative¹⁷).

En este sentido, la IFLA ha centrado sus esfuerzos en desarrollar un modelo conceptual que sirva de referencia para la representación del universo bibliográfico. El modelo desarrollado por el Working Group on Functional Requirements for Bibliographic Records se basa en las entidades asociadas a los registros bibliográficos y en los diferentes tipos de relaciones entre estas entidades y el propio registro (*Functional requirements for bibliographic records*, 1998).

Con el mismo objetivo y siguiendo los mismos criterios se creó el Working Group on Minimal Level Authority Records. Su finalidad era diseñar un formato de autoridades

¹⁶ Dublin Core Metadata Initiative <http://uk.dublincore.org/> [Consulta 20 de marzo de 2003]

¹⁷ Text Encoding Initiative <http://www.tei-c.org/> [Consulta 20 de marzo de 2003]

flexible y adaptable a diferentes tipos de información, para lo que se identificaron los elementos esenciales de un registro de autoridad y se definió cada uno de ellos. Tras la publicación del informe *Mandatory Data Elements for Internationally Shared Resource Authority Records* (1998) la investigación continúa a cargo del grupo de trabajo FRANAR (Funcional Requirements and Numering Authority Records) que también analiza la viabilidad de crear un número normalizado para registros de autoridad (ISAND: International Standard Authority Data Number) con la finalidad de facilitar el intercambio de registros (Patton 2003).

Otro núcleo de estudios es el que se centra en la definición de formatos alternativos al MARC para la codificación de la información bibliográfica. En este caso la mayoría de los grupos trabaja en el desarrollo de programas que permitan convertir datos MARC a SGML/XML. Algunos ejemplos los constituyen los trabajos de UKOLN¹⁸ (United Kingdom Office for Library Information Networking de la University of Bath), W3C¹⁹ (WWW Consortium), la Library of Congress o el Dublin Core²⁰. Este tipo de formatos presenta varias ventajas (Gaynor 1996; Ortíz-Repiso 1998; Ortiz Repiso y Moscoso 1999):

¹⁸ UKOLN <http://www.ukoln.ac.uk/> [Consulta 20 de marzo de 2003]

¹⁹ W3C Metadata and resource description: <http://www.w3.org/Metadata/> [Consulta 20 de marzo de 2003]

²⁰ Se puede encontrar una relación de los proyectos en: http://orc.rsch.oclc.org:6464/related_projects.html [Consulta 20 de marzo de 2003]

- Permite indizar y recuperar tanto los documentos como los componentes de los documentos, lo que facilita el desarrollo de un sistema de nodos y enlaces propio del entorno Web.

- La estructura textual de SGML o XML soporta una navegación avanzada en línea que permite que se maneje la información utilizando relaciones jerárquicas a tantos niveles como sea necesario.

- También permite una flexibilidad máxima en el uso del texto: el usuario puede controlar los formatos de indización, visualización e impresión y los registros se pueden interrelacionar en diferentes ficheros. La utilización de las DTD (Document Type Definition) y hojas de estilo permiten flexibilizar al máximo la explotación y utilización de los datos.

A mediados de los noventa, la Network Development and MARC Standard Office de la Library of Congress puso en marcha el proyecto MARC DTD (Machine Readable Cataloging Document Type Definition)²¹. Las dos principales realizaciones del proyecto fueron elaborar las DTD's de SGML correspondientes a los cinco formatos

²¹ Marc in XML <http://www.loc.gov/marc/marxml.html> [Consulta 20 de marzo de 2003]

MARC²² y desarrollar los programas informáticos para realizar las conversiones entre ambas normas de codificación. (García Camarero y García Melero 2001)

Uno de los productos mejor acabados que ha desarrollado la Library of Congress es el MARC21 XML Schema²³. Dentro de este marco no sólo propone una estructura genérica y se facilitan herramientas para compatibilizar los dos formatos, sino que también se ofrece la posibilidad de transformar la información en metadatos del esquema Dublin Core. De hecho, finalmente se ha conseguido que los datos de estos dos formatos sean totalmente intercambiables como se ha demostrado en el proyecto CORC (actualmente CONNEXION²⁴) donde se han creado y gestionado simultáneamente registros bibliográficos en MARC 21 y Dublin Core.

1.3.2 Contenido informativo del registro bibliográfico.

El contenido informativo de los registros de los catálogos en línea ha sido objeto de numerosas críticas porque, al ser muy reducido, limita el número de puntos de acceso y dificulta los juicios sobre la relevancia de los registros recuperados (Belkin y Saracevic

²² Los cinco formatos se agruparon en dos DTD's: bibliográfica y de autoridades. La primera comprende los formatos para registros bibliográficos, de fondos y de información a la comunidad. La segunda los de autoridades y clasificación bibliográfica. Posteriormente las dos DTD's se actualizaron para XML.

²³ MarcXml: MARC 21 XML Schema oficial web site. <http://www.loc.gov/standards/marcxml/> [Consulta 20 de marzo de 2003]

²⁴ <http://connexion.oclc.org/> [Consulta 20 de marzo de 2003]

1992, O'Brien 1994). Esta apreciación afecta especialmente a la información temática, que se considera no sólo escasa sino poco apropiada, ya que los tradicionales encabezamientos de materia presentan problemas de inconsistencia, carencia de especificidad y dificultad para ser entendidos por usuarios no expertos.

El incremento de información temática relativa a las monografías comenzó a practicarse en la década de los setenta (Fernández Molina y Moya Anegón 1998 Fernández Molina y Peis 1995). El proyecto pionero en este tema fue el Subject Access Project (SAP), dirigido por Atherton (Atherton 1978) en el que se incorporaron a los registros términos del índice y de las tablas de contenido. Se puede considerar que éste fue el punto de partida de otros proyectos como el SAP-Sweden (Wormell 1985), el Engineering Information System (EIS) de la Universidad de Purdue (Posey y Erdmann 1986), el ESP (Enhancing Subject Project) en la Australian Defence Force Academy (Byrne y Micco 1988). En estas y en las investigaciones de Hart y Reitsma (1990), Markey (1987a,1989), Michalak (1990) o Peis (2000) se han empleado diversos medios para enriquecer los registros con más información sobre materia:

- Se incluye la tabla de contenido del documento en el campo notas y se permite la búsqueda por palabras clave en este campo.

- Se añaden al registro términos seleccionados de los índices del libro y se permite la búsqueda por palabras clave sobre estos términos.

- Se realiza una indización más exhaustiva de cada documento, asignándole más descriptores.

- Se incluyen en el registro los términos de materia que corresponden a los códigos de clasificación en las tablas y se permite la búsqueda por palabras clave sobre estos términos.

Las investigaciones han demostrado que el enriquecimiento de los registros por alguno de estos medios incrementa la circulación de los documentos y mejora el acceso temático. La probabilidad de recuperar registros relevantes aumenta notablemente y, consecuentemente, se reduce el número de búsquedas que no recuperan nada. Por ejemplo, en el estudio de Byrne y Micco (1988) se estimó que la recuperación se incrementaba aproximadamente en un 300%. También permite al usuario utilizar más términos y más específicos que si sólo se utilizaran encabezamientos de materia. El análisis realizado por Peis (2000) sobre el incremento de información que supone la incorporación de las tablas de contenido de monografías colectivas demostró que el 95,7% de los títulos individuales permiten identificar el contenido y que añadirlos al registro incrementa el número de elementos temáticos en un 700%.

La desventaja es que también se recuperan registros no relevantes y, además, hay muchas más posibilidades de que se recuperen muchos registros que el usuario tendrá que visualizar (Byrne y Micco 1988; Lancaster et al. 1991). Esto puede suponer un serio obstáculo porque el usuario debe conocer cómo concretar la búsqueda cuando se han recuperado demasiados registros.

Otro problema añadido es que el enriquecimiento supone costes adicionales en términos de recursos económicos y humanos, espacio de memoria para el almacenamiento de información y tiempo de procesamiento. La utilización de las nuevas tecnologías puede disminuir la incidencia de estos problemas.

Un ejemplo es el proyecto RIDDLE (Rapid Information Display and Dissemination in a Library Environment), financiado por el Programa de Bibliotecas de la Comisión de las Comunidades Europeas y desarrollado entre 1994 y 1995 (Harrison, Roos y Thomas 1995) o la investigación que, utilizando una metodología similar, llevaron a cabo Peis y Fernández Molina (1998). En ambos trabajos se estudió la viabilidad del uso de la tecnología del escáner para capturar las tablas de contenido de publicaciones periódicas y de monografías colectivas. A continuación se extrajo de ellas información textual utilizando programas de reconocimiento óptico de caracteres (OCR) y se cargaron estos datos en un catálogo en línea etiquetándolos con SGML. Se hizo

especial hincapié en evaluar el coste del sistema en comparación con otros métodos. Aunque los resultados de las pruebas fueron satisfactorios en términos generales, también demostraron que la mejora de la base de datos tiene que complementarse con el desarrollo de mecanismos de búsqueda que permitan contrarrestar el problema de la sobrecarga de información.

1.3.3 Técnicas de recuperación.

La evolución de las técnicas de recuperación implementadas en los OPAC sin duda ha contribuido a mejorar el acceso por materias y a superar algunos de sus problemas. Por ejemplo, la opción de buscar por palabras clave junto con la eliminación de las palabras vacías reduce los problemas que supone la sintaxis de los encabezamientos porque esta manera sólo se tiene en cuenta si están o no los términos significativos, pero no el orden en que aparezcan. Otras alternativas, como el truncamiento o la búsqueda por palabras clave en el registro, amplían el ámbito de la búsqueda temática, incrementan el vocabulario de entrada y reducen así las posibilidades de que se produzca silencio documental.

Sin embargo, estas técnicas se han mantenido dentro de un paradigma tradicional, orientado a las búsquedas analíticas basadas en la simple equiparación del enunciado de

búsqueda con la representación de los documentos almacenados en el sistema. Además, dentro de este modelo, han prevalecido las técnicas de equiparación exacta que utilizan la lógica booleana y los operadores de proximidad para indicar la relación entre los términos, junto con la opción de restringir la búsqueda a campos concretos (Hildreth 1995a)

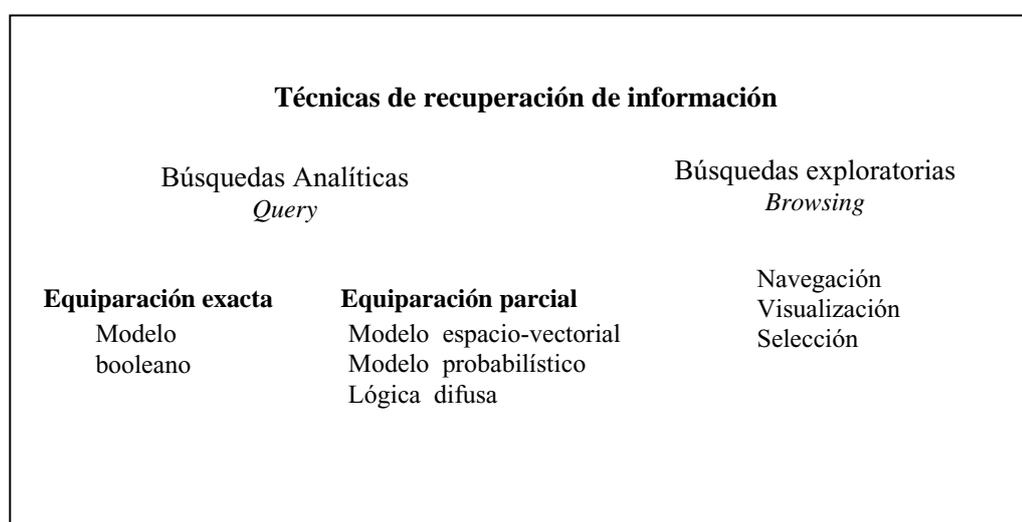


Fig. 1.11 Técnicas de recuperación de información (Adaptado de Hildreth 1995a)

Aunque la combinación de estas opciones permite elaborar búsquedas muy sofisticadas, también plantea muchos problemas (Doszkocs 1983; Mischo y Lee 1987; Muñoz García1995; Porter y Galpin 1988; Salton, Buckley y Fox 1983). El modelo booleano posee limitaciones importantes que han suscitado numerosas críticas:

- √ La lógica booleana no se adquiere por intuición, sino que requiere formación. Como consecuencia, la mayor parte de los usuarios no son capaces de generar una ecuación de búsqueda eficaz sin la ayuda de un profesional.

- √ Por otro lado, en la respuesta a una pregunta, el álgebra de Boole no es capaz de mostrar los documentos en función de algún grado de relevancia: el documento es recuperado (1) si satisface la ecuación de búsqueda o no (0) si no la satisface.

- √ En estas condiciones el operador AND puede resultar demasiado restrictivo y provocar silencio documental. Al contrario, el operador OR es muy expansivo y por tanto puede producir ruido.

- √ Además, dado un par de términos sobre los que se quiere consultar, al usuario no le es posible establecer si el término X es más importante que el término Y ni en qué proporción.

- √ El resultado de una búsqueda muchas veces es una enorme cantidad de documentos sin ningún orden de relevancia. Por ejemplo, ante una pregunta que utilice varios términos unidos por un OR (*A or B or ... Z*) se recupera con la misma relevancia un documento que tenga sólo un término que otro que los

tenga todos. Sería deseable que se mostraran en primer lugar aquellos documentos con los todos los términos, posteriormente aquellos con todos menos uno, etc.²⁵.

Para evitar los problemas que se derivan de estas limitaciones se han desarrollado nuevas técnicas de recuperación de información. En el caso de los catálogos en línea, las soluciones que se han adoptado tienen dos orientaciones diferentes: simplificar el uso de las técnicas tradicionales y utilizar técnicas avanzadas de recuperación de información.

1.3.3.1 Simplificación de las técnicas tradicionales.

En este caso se trata de incorporar al sistema de búsqueda algún mecanismo que ejecute, de forma automática y transparente, opciones de búsqueda con las que los usuarios normalmente no tienen experiencia. Por ejemplo, una alternativa es la formulación implícita de la búsqueda en la que el sistema especifica automáticamente las relaciones lógicas entre los términos utilizados por el usuario: se puede incluir un truncamiento, un Y lógico entre los términos, etc. (Hildreth 1995a).

²⁵ En esta idea se basa el modelo booleano extendido.

Otra posibilidad es emplear secuencias de búsqueda preestablecidas que se ejecutan de forma rutinaria en función de algún parámetro preestablecido. Cheng (1985) describe una posible secuencia de estrategias para la búsqueda por materias: incorporación de truncamientos, presentación alfabética de encabezamientos similares, separación de los términos de búsqueda utilizando un Y booleano y finalmente búsqueda de todas las palabras en el título. Si la búsqueda por título recupera algún registro relevante, el sistema utiliza el primer encabezamiento de materia del registro y continúa la búsqueda por materias con ese encabezamiento. Otra muestra de esta solución son los “árboles de búsqueda” (Drabenstott y Vizine-Goetz 1994; Drabenstott 1996a) mediante los cuales el sistema decide el tipo de búsqueda más adecuada en función de las características del enunciado utilizado por el usuario y de los resultados que se van obteniendo: búsqueda en orden directo, truncamiento a la derecha, presentación del enunciado en el índice alfabético, palabras clave en el campo materias y palabras clave en el registro.

1.3.3.2 Técnicas avanzadas de recuperación de información.

La segunda opción es utilizar técnicas avanzadas de recuperación de información, basadas en la equiparación parcial. Con estas técnicas se calcula la similitud entre el enunciado de búsqueda y los documentos de la base de datos (o sus representaciones).

Los documentos recuperados se parecerán más o menos al enunciado de búsqueda pero no tienen por qué cumplir todas las condiciones. Los resultados se ordenan en función de la similitud, de manera que los documentos más similares aparecerán en los primeros lugares de la lista.

Se han realizado múltiples clasificaciones de estas técnicas (Belkin y Croft 1987; Frakes y Baeza-Yates 1992; Muñoz García 1995; Rijsbergen 1979; Salton y McGill 1983; Sparck Jones y Willet 1997), aunque todas coinciden en señalar tres grupos fundamentales basados en los modelos formales que las sustentan:

- ✓ Modelo probabilístico, basado en enfoques bayesianos con los que se construyen espacios de sucesos y funciones de probabilidad de documentos y características. Las funciones de probabilidad extraen aquellos documentos con más probabilidad de responder a la búsqueda planteada.

- ✓ Las técnicas basadas en la lógica difusa se basan en la modificación del concepto de pertenencia a un conjunto: en la lógica booleana un elemento pertenece o no a un conjunto mientras que en la lógica difusa hay un grado de pertenencia continua que va desde 0 a 1. Cabe considerarlas un caso particular de las técnicas probabilísticas porque para construir las funciones concretas de pertenencia a un conjunto hay que basarse en estadísticas o

ejemplos concretos, lo que asume implícitamente un modelo probabilístico (Muñoz García 1995).

- ✓ Modelo de espacio vectorial, en el que documentos y búsquedas se expresan como combinaciones lineales en un espacio vectorial de n-dimensiones (tantas como términos de búsqueda se utilicen en la base de datos). La función de recuperación suele ser una función de distancia entre preguntas y documentos.

Estas técnicas de recuperación se implementan normalmente en sistemas en los que se extraen automáticamente los términos que servirán como características para representar los documentos. Para evitar que con los datos así obtenidos se generen enormes ficheros de búsqueda se utilizan distintos sistemas de “filtrado” que permiten reducir el tamaño de los ficheros y valorar, en alguna medida, la capacidad de los términos para representar el contenido de los documentos (Baeza-Yates y Ribero-Neto 1999; Harman 1994; Moreiro González 2002).

La forma más habitual de eliminar los términos no significativos es utilizar un antídiconario o lista de palabras vacías. Sin embargo, muchos sistemas de recuperación de información incluyen mecanismos que permiten reducir el número de términos de indización mediante algún tipo de control de las formas flexionadas y

derivadas. Si se considera que aquellos términos que tienen la misma raíz tienen también un significado equivalente, la *reducción de términos a la raíz* (lematización²⁶) servirá para reducir considerablemente el fichero de búsqueda sin que eso implique una pérdida sustancial de información.

Los algoritmos de lematización eliminan de forma automática sufijos y prefijos (en algunos casos también interfijos). Muchas veces con esta eliminación se obtiene como resultado la raíz, otras sólo se reducen las flexiones de una forma gramatical y otras simplemente se produce un truncamiento de las palabras. Frakes (1992) clasifica los métodos automáticos para la lematización en:

- *Algoritmos de eliminación de afijos*: eliminan sufijos o prefijos del término y dejan un lexema siempre que éste cumpla las condiciones impuestas en las tablas del algoritmo. Se han desarrollado con éxito para la lengua inglesa (Porter 1980), pero no funcionan bien para lenguas con flexiones más irregulares y ricas como el español (Honrado et al. 2002).
- *Búsqueda en tabla*. Los términos y sus correspondientes lexemas pueden estar almacenados en una tabla realizando la lematización a través de búsquedas en ella.

²⁶ En adelante utilizaremos el término *lematización* que es la traducción más frecuente al español del término inglés *stemming*.

- *Variedad de sucesores*: Están basados en el cálculo de la longitud de los prefijos que mejor admiten la expansión mediante la incorporación de sufijos.
- *N-gramas*. En esta aproximación se calculan medidas de asociación entre pares de términos basándose en los N-gramas²⁷ únicos compartidos. Los términos que superan determinado umbral de similitud se consideran equivalentes.

Otro proceso asociado a las técnicas avanzadas de recuperación de información es la ponderación de los términos de indización, que consiste en asignar un valor a los términos extraídos de los documentos. Basándose en los postulados de Luhn (1957), el número de veces que aparecen los términos en los documentos se utiliza como criterio para valorar la capacidad que éstos tienen para representar la información. Los algoritmos de ponderación más utilizados asignan pesos que son directamente proporcionales a la frecuencia de aparición del término en el documento y que están inversamente relacionados con el número de veces que éste aparece en la base de datos. Puede considerarse un método de “filtrado” si se eliminan aquellos términos que no alcanzan determinado peso, aunque habitualmente se utiliza para afinar el cálculo de la similitud búsqueda/documento y mejorar así el ranking de salida (Harman 1992).

Aunque no se trata exactamente de una técnica de recuperación, otra buena herramienta para evitar el silencio documental en los sistemas de recuperación de información es la detección y corrección automática de errores tipográficos y ortográficos. Sus efectos resultan tan evidentes que es difícil entender porqué no se implementa habitualmente en los catálogos en línea aunque sí se ha hecho de manera experimental.

Por ejemplo, Walker y Jones (1987) compararon dos versiones de Okapi con dos soluciones diferentes para asistir al usuario en la detección de errores de este tipo. En la primera versión los errores potenciales se detectaron utilizando el “Soundex algorithm”, se informaba al usuario de la posibilidad de que la palabra estuviera mal escrita y se sugería una palabra alternativa. La segunda versión era similar, pero en este caso se proponía al usuario que utilizara un término nuevo para corregir el potencial error detectado por el algoritmo sin sugerirle ninguno. Los resultados demostraron que el primer sistema mejoraba los resultados en un 78% de los casos mientras que el segundo lo hacía en un 64%.

El principal foro para la investigación relacionada con las técnicas de recuperación de información son las Text REtrieval Conferences (TREC 1992-) coordinadas por el

²⁷ Secuencias de caracteres de longitud previamente establecida: dos caracteres los digramas, tres los trigamas, etc.

National Institute of Standards and Technology (NIST). Desde 1992, en estos experimentos, equipos de investigación de diferentes países prueban sus sistemas de recuperación de información realizando búsquedas contra una colección experimental en la que previamente se han establecido juicios sobre la relevancia de los documentos (Harter 1997; Olvera Lobo 1999). A lo largo de estos años, en este marco se ha evaluado la eficacia de multitud de algoritmos de lematización y ponderación, funciones de similitud y modelos de representación y recuperación. Además, desde su tercera edición incluyen una “*Interactive Track*” en la que se evalúa la incidencia que tiene la implementación de estas técnicas en la interacción usuario-sistema (Over 2001).

Algunos autores han abogado por aplicar en los OPAC’s las técnicas de equiparación parcial contrastadas en estas conferencias. Si se emplearan se recuperarían aquellos registros que contengan todos o alguno de los términos de búsqueda utilizados por los usuarios (o su forma lematizada), ordenados en función de su relevancia respecto a la búsqueda. Sin embargo, esta alternativa plantea algunos problemas:

- ✓ En primer lugar, son técnicas pensadas para sistemas que trabajan con más información que la que contiene un registro bibliográfico. Los registros de un catálogo en línea son menos ricos en información que los de otros

sistemas de recuperación de información, de forma que los índices que se crean a partir de los datos que contienen también son más limitados.

- ✓ Por otra parte, los mejores resultados con estas técnicas se obtienen cuando se emplean enunciados de búsqueda largos. Este tipo de planteamientos no es habitual en un catálogo en línea donde los usuarios, como ya hemos visto, tienden a utilizar expresiones breves. En estas condiciones, aunque su uso reduzca el silencio documental, incrementa peligrosamente el riesgo de sobrecarga de información.

A pesar de estos inconvenientes, las investigaciones han demostrado que su utilización en los catálogos en línea puede mejorar los resultados de la búsqueda por materias. La mayoría de estos proyectos se han desarrollado sobre catálogos experimentales de los cuales los más relevantes son²⁸:

- Okapi (Robertson 1997, Robertson, Walker y Beaulieu; 2000 Walker 1989) es un proyecto que se inició en 1982 en la Polytechnic of Central London y que, desde 1989, continúa en el Centre for Interactive Systems Research at the City University²⁹. A lo largo de estos años se han desarrollado diferentes versiones del

²⁸ Puede verse una relación de los proyectos en Onion Patch(sm): New age public access systems: projects, research, products and services. <http://www.public.iastate.edu/~CYBERSTACKS/Onion.htm> [Consulta 27 de marzo de 2003]

²⁹ <http://www soi.city.ac.uk/organisation/is/research/cisr/> [Consulta 27 de marzo de 2003]

catálogo experimental con las que se ha investigado el efecto de factores como: niveles de lematización, utilización de referencias cruzadas automáticas, corrección automática de errores ortográficos, técnicas de recuperación probabilísticas y ordenación por relevancia de los resultados. Numerosos trabajos se han centrado en comprobar el efecto que produce la implementación de estas técnicas sobre los resultados de las búsquedas, especialmente de materias. Sin embargo, el interés fundamental ha sido buscar métodos para mejorar la interacción con los usuarios durante la búsqueda de información.

- Otro de los proyectos pioneros fue CITE (Doszkocs 1983), un *front-end*³⁰ para catálogo en línea de la National Library of Medicine (CATLINE). CITE lematizaba las palabras clave de las búsquedas de los usuarios e identificaba variantes de las mismas. Posteriormente ponderaba estos términos utilizando algoritmos de frecuencia inversa, aunque la ponderación podía ser modificada por el usuario, a quien se daba la opción de reordenar los términos. A continuación, se sumaba el peso de los términos que aparecían en los registros recuperados y se ordenan los resultados en función de este valor. Las investigaciones continuaron con The Library of Congress Experimental Search System (Jacso 1998; Wilder y Greenfield 1997) que facilitaba acceso tanto a registros bibliográficos como a recursos

³⁰ Interfaz entre el usuario y la base de datos que posibilita la consulta y facilita algunos pasos del proceso como la formulación pre-búsqueda, la descarga de información, etc. (Eftimiadis 1990)

digitales. En este caso se utilizó el sistema de recuperación INQUERY³¹, basado en técnicas probabilísticas, tanto para ponderar los términos como para ordenar los resultados de búsqueda por relevancia.

- Chesire II³², desarrollado en la School of Information Management and Systems at the University of California, Berkeley (Larson 1991a, 1992; Larson et al. 1996), es otro ejemplo de catálogo experimental que pretende servir de puente entre los sistemas puramente bibliográficos y los de texto completo. El sistema combina las técnicas de recuperación probabilísticas con un sistema basado en SGML que utiliza el protocolo Z39.50. También emplea el modelo probabilístico para clasificar automáticamente los documentos (clustering). La combinación de estas técnicas mejora los resultados, alcanzando los niveles logrados por un usuario experimentado, familiarizado con los términos de la LCSH, la lógica booleana, el truncamiento y la exploración de la base de datos (Larson et al. 1996).
- Oasis³³ (Otlet's Adaptive Search Information Service) es un proyecto de investigación de la University of California at Berkeley en el marco del Digital Library Project financiado por la National Science Foundation (NSF), Defense

³¹ Sistema desarrollado por el Centre for Intelligent Information retrieval at the University of Massachusetts, Amherst (<http://ciir.cs.umass.edu/index.html>) [Consulta 27 de marzo de 2003]

³² <http://elib.cs.berkeley.edu/cheshire> [Consulta 27 de marzo de 2003]

³³ <http://www.sims.berkeley.edu/research/projects/oasis/> [Consulta 27 de marzo de 2003]

Advanced Research Projects Agency (DARPA) y National Aeronautics and Space Administration (NASA). En principio, las investigaciones se centraron en crear un prototipo de *front-end* que resolviera los problemas que provocaba la utilización de la lógica booleana en el catálogo en línea MELVYL. Se desarrollaron aplicaciones para manejar los resultados de las búsquedas que recuperaban demasiados registros mediante procedimientos de partición y ordenación por relevancia. Actualmente, los trabajos se dirigen a facilitar la gestión de las búsquedas y la manipulación de los resultados cuando se consultan simultáneamente varios catálogos en línea y a aplicar técnicas de procesamiento del lenguaje natural a la recuperación de información (Buckland 1995; 2000; Buckland et al.1992).

- MARIAN³⁴ (Fox et al 1993; Gonçalves, France y Fox 2001) es un sistema de indización y recuperación optimizado para bibliotecas virtuales, desarrollado por el Virginia Tech Computing Center for VT Information System. Sus características se basan en los resultados del Retrieval Experiment – Virginia Tech Online Catalog (REVTOLC) en el que se implementaron técnicas de recuperación basadas en el modelo vectorial para buscar en un catálogo experimental con registros en formato MARC. Sobre esta base, MARIAN se desarrolla como un sistema de búsqueda en un catálogo colectivo virtual en el que

³⁴ <http://www.dlib.vt.edu/products/marian.html> [Consulta 27 de marzo de 2003]

el reto es integrar el acceso a objetos de información diferentes, con las mismas técnicas de búsqueda y desde una misma interfaz.

Como consecuencia de los buenos resultados obtenidos experimentalmente, el uso de estas técnicas está empezando a extenderse a los catálogos comerciales (Jones, Cunningham, McNab, R. 1998), aunque no de manera generalizada y muchas veces como módulos “adquiribles” al margen de las opciones standard. Por ejemplo, empieza a ser frecuente que se incorpore algún mecanismo que controle las diferencias singular/plural como sucede en el catálogo colectivo del Consortium of University Research Libraries (COPAC)³⁵. Oracle⁷³⁶, sistema de recuperación de información desarrollado por Fretwell-Downing Informatics, incluye técnicas avanzadas de recuperación de información contrastadas previamente en Okapi: lematización y ponderación de los términos, equiparación parcial y ordenación por relevancia (Joy 1995).

Cada vez es más frecuente encontrar también la opción de ordenar los resultados de acuerdo a algún criterio de relevancia, como sucede en la búsqueda por palabras clave de INNOPAC/Millennium³⁷ (Innovative Interfaces) (Greene 1996) en el OPAC de la

³⁵ <http://copac.ac.uk/copac/> [Consulta 27 de marzo de 2003]

³⁶ Puede consultarse por ejemplo en la biblioteca del British Geological Survey <http://geolib.bgs.ac.uk/> [Consulta 27 de marzo de 2003]

³⁷ Esta opción está implementada en el catálogo CISNE de la Universidad Complutense de Madrid <http://cisne.sim.ucm.es> [Consulta 27 de marzo de 2003]

Library of Congress³⁸. En este último caso el algoritmo de relevancia que se emplea para ordenar los documentos tiene en cuenta los siguientes factores para ponderar los términos:

- su frecuencia de aparición en la base de datos,
- su proximidad dentro del registro,
- el número de veces que aparecen dentro de un campo o dentro del registro
- que el usuario los haya marcado como términos imprescindibles (+) o como términos muy relevantes (*)

1.3.3.3 Modificación automática de la búsqueda.

Las técnicas que hemos visto hasta ahora facilitan la equiparación entre las representaciones de los documentos y las de las búsquedas pero no solucionan uno de los mayores problemas de los usuarios: ¿cómo modificar una búsqueda cuando los resultados no son satisfactorios? En esta circunstancia, los especialistas buscan términos alternativos en los vocabularios controlados o utilizan descriptores y terminología que aparece en documentos que previamente han identificado como relevantes para la materia. Sin embargo, los usuarios no expertos no poseen

³⁸ <http://catalog.loc.gov> [Consulta 27 de marzo de 2003]

normalmente los recursos necesarios para realizar este proceso por lo que sería interesante que el sistema lo ejecutara de forma automática.

La operación de modificación o ampliación de la búsqueda puede realizarse de forma automática tomando los términos de las estructuras de conocimiento o de los documentos recuperados. En la primera opción se incluyen en el enunciado los términos relacionados procedentes de algún tipo de estructura de conocimiento. La segunda posibilidad utiliza las técnicas avanzadas de recuperación para realizar una retroalimentación por relevancia. En este proceso, los documentos considerados relevantes en una iteración previa se convierten en fuente de términos de búsqueda en la reformulación. Esta definición corresponde a la retroalimentación por relevancia positiva, la más habitual en los sistemas de recuperación de información. Sin embargo, en algunos sistemas experimentales se trabaja también con la retroalimentación por relevancia negativa: se excluyen de la búsqueda los términos de aquellos documentos que no se han considerado relevantes (Efthimiadis 1996; Harman 1992a; Spink y Loose 1996).

Esta opción puede implementarse perfectamente en un catálogo en línea para reformular la búsqueda del usuario y para depurar los resultados. El usuario puede indicar cuál o cuáles de los registros recuperados son relevantes, y una rutina automática se encarga de recuperar otros similares a los que se han marcado porque

comparten los mismos encabezamientos de materia, las mismas palabras en el título, etc. (Hildreth 1995a; Beheshti 1997). De hecho, empieza a incorporarse a algunos sistemas comerciales: por ejemplo, INNOPAC/Millennium (Innovative Interfaces) ofrece la posibilidad de “Mostrar ejemplares similares” en la búsqueda por palabras clave. Si el usuario selecciona esta opción, el sistema utiliza los encabezamientos de los registros seleccionados para realizar una búsqueda por palabras clave en el campo materia.

Experimentalmente se ha utilizado en muchos de los proyectos que ya hemos reseñado como herramienta para mejorar la interacción usuario-sistema. Por ejemplo, el sistema CITE (Doszckocs 1983) realizaba un análisis de la frecuencia de los encabezamientos de materia de los registros que el usuario había considerado relevantes y le presentaba una lista para que seleccionara los más adecuados. Estos encabezamientos de materia junto con los códigos de clasificación de los registros seleccionados se utilizaban en una búsqueda posterior por equiparación parcial.

En el marco del proyecto Okapi se ha investigado la eficacia de diferentes algoritmos de retroalimentación y de distintas presentaciones, tanto para mejorar los resultados como para facilitar la interacción (Beaulieu 2001). Las investigaciones sobre la “ampliación de la búsqueda” se iniciaron en la versión Okapi3 que utilizaba como fuentes de términos el título, los encabezamientos y los códigos de clasificación de los

registros considerados relevantes por los usuarios. A continuación, se les que asignaba un peso en función del número de registros seleccionados en los que aparecieran y se seleccionaban los 32 términos más relevantes (Walker y De Vere 1990).

En versiones posteriores se ha evaluado si los resultados son mejores cuando se retroalimenta la búsqueda automáticamente o cuando se hace de forma interactiva. En el segundo caso se le presenta al usuario una lista de términos - previamente seleccionados por el sistema después de ponderar la terminología de los registros relevantes - para que él seleccione los que quiere incluir en la modificación. También se investigado como incide en el proceso que la interacción se realice con una interfaz basada en menús de texto o con una gráfica (XOkapi). Los resultados parecen demostrar que resulta más eficaz la retroalimentación de forma interactiva realizada con una interfaz gráfica (Beaulieu 1997; Hancock-Beaulieu, Fieldhouse y Do 1995).

1.3.4 Estructuras de conocimiento.

A lo largo de las dos últimas décadas, las investigaciones han demostrado que los usuarios tienen problemas tanto para encontrar los términos que representen su necesidad de información como para modificarlos cuando los resultados no son satisfactorios. Además, como señalan Sonnenwald e Iivonen (1999), la necesidad de

información es una entidad dinámica que puede cambiar durante el proceso de interacción con el sistema de recuperación. Como consecuencia de este carácter, durante el desarrollo de una búsqueda el usuario muchas veces cambia de idea sobre lo que prefiere y necesita modificar la manera de expresar esta necesidad. Para facilitar estas búsquedas, en las que muchas veces “no se sabe lo que se quiere hasta que se encuentra” (Cove y Walsh 1988), es importante mostrar al usuario cómo está organizada la información, qué conceptos se han empleado para caracterizarla y las conexiones que existen entre ellos.

Durante los noventa, la evolución de los OPAC's se centró en incorporar atractivas técnicas de recuperación que parecían el medio adecuado para solucionar todos los problemas: búsqueda por palabras clave, equiparación parcial, ordenación por relevancia de los resultados, etc. Sin duda estas técnicas han supuesto una importante mejora para los catálogos en línea. Sin embargo, el énfasis que se ha puesto en ellas ha ido en detrimento de otras opciones que favorecen el acceso mediante búsquedas exploratorias (Greenberg 1997).

En el contexto de este último tipo de búsquedas, los sistemas para la organización del conocimiento³⁹ tradicionalmente utilizados en las unidades de información

³⁹ Se utilizarán las expresiones “sistemas de organización del conocimiento” y “estructuras de conocimiento” indistintamente para referirnos a los sistemas empleados en entornos manuales o automatizados para representar y organizar la información contenida en los documentos y, por extensión, la misma documentación.

(clasificaciones, encabezamientos de materia y tesauros) han dejado de considerarse el problema que parecían ser para los defensores de sistemas automatizados con indización en lenguaje natural (Bates 1998; Chan 2000). Por el contrario, han pasado a tratarse como herramientas eficaces para la búsqueda, combinándolas con técnicas avanzadas de recuperación de información y con el hipertexto para facilitar la navegación. Como sugiere López-Huertas (1997), refiriéndose concretamente a los tesauros, pueden jugar un papel de intermediarias entre el usuario y el sistema en la medida en que representan tanto las estructuras de conocimiento transmitidas por los creadores de la información, como la estructura cognitiva de los usuarios tal como se manifiesta en las búsquedas.

Por otra parte, el desarrollo de la biblioteca digital está actualizando necesidades patentes durante años. Resulta evidente que la tecnología permite ya, o permitirá a corto plazo, buscar en diferentes depósitos de objetos de información, manejando una amplia variedad de protocolos y formatos, mediante algún tipo de software que compense las diferencias entre sistemas (Paepcke 1996; *Workshop on Distributed Heterogeneous Knowledge Networks* 1997). Sin embargo, el objetivo final de la biblioteca digital, como señalaba el informe Information Infrastructure Technology and Applications (IITA 1995), es que la conexión de sistemas sea operativa a nivel semántico. Es decir, conseguir que un usuario pueda acceder de forma consistente y coherente a clases similares de objetos y servicios distribuidos en depósitos

heterogéneos.(Chen 1999; Dahlberg 1996; Hodge 2000). Las líneas de trabajo en este sentido incluyen la definición y utilización de metadatos, el uso de descripciones de los objetos en formatos informáticos normalizados, la integración de depósitos heterogéneos con semántica diferente y la organización temática de la información.

La atención a la interoperabilidad semántica ha dado lugar a diferentes proyectos de investigación a cargo de la Digital Libraries Initiative⁴⁰, promovida por NSF, DARPA y NASA, o del Information Society Technologies Programme⁴¹ de la Comisión Europea. Entre las investigaciones más relevantes en el terreno de las técnicas semánticas de análisis, indización y sistematización destacan las orientadas a:

- La utilización de técnicas de indización basadas en el procesamiento del lenguaje natural. Se aplican métodos de análisis morfosintáctico y categorización semántica para obtener palabras o frases que permitan caracterizar los documentos. El objetivo de la incorporación de estas técnicas es optimizar los resultados, evitando tanto los problemas que plantea la ambigüedad del lenguaje natural como las restricciones de los vocabularios controlados (Doszkocs 1987; Cheng y Wilensky 1997; Gonzalo, Verdejo y Peña 2002; Moreiro González y Méndez Rodríguez 1999; Pérez-Carballo y Strzalkowski 2000; Strzalkowski et al.1999; Wilensky 1996). Sin embargo, la

⁴⁰ <http://www.si.umich.edu/SantaFe/> [Consulta 6 de abril de 2003]

⁴¹ <http://www.cordis.lu/ist/ist-fp5.html> [Consulta 6 de abril de 2003]

experimentación con estos métodos no ha mejorado sustancialmente los resultados de búsqueda (Sparck Jones 1999). Ante esta evidencia, las investigaciones se han dirigido a complementarlos con redes semánticas, como WordNet o EurowordNet (Vossen 1998), y tesauros.

- El análisis y organización semántica de textos y objetos multimedia. Las estructuras de conocimiento que se han utilizado con esta finalidad pueden clasificarse en dos grandes grupos: dependientes de la colección e independientes de la colección.
 - √ Las estructuras de conocimiento dependientes de la colección se generan a partir de las características de los documentos almacenados en el sistema. Las características (términos, citas, enlaces, etc.) sirven como criterios para crear agrupaciones mediante técnicas de clustering (heurísticas o jerárquicas), métodos estadísticos multivariable (análisis factorial, escalamiento multidimensional), redes neuronales, programación evolutiva y algoritmos genéticos (Baeza-Yates y Ribero-Neto 1999; Chen 1995; Kohonen et al. 1999).

Estas técnicas pueden utilizarse para crear grupos de documentos generando clasificaciones automatizadas (Gerrero, Moya-Anegón y

Herrero-Solana 2002; Johnson y Fotouhi 1998; Larson 1996; MacKiernan 2002; Schatz 1999). También pueden servir para crear agrupaciones de términos dando lugar a “tesauros por asociación”, basados en la co-ocurrencia de las palabras, o a “tesauros por similaridad” en los que cada término se convierte en un vector cuyas dimensiones son los documentos de la base de datos (Crouch y Yang 1992; Guerrero et al. 2002; Jing, Croft y Bruce 1994)

- √ Las estructuras de conocimiento independientes de la colección son sistemas que permiten la organización de un área temática, general o especializada, independientemente de los documentos almacenados la base de datos. Se trata de las clasificaciones y los vocabularios controlados empleados tradicionalmente en catálogos en línea y bases de datos bibliográficas. A ellos se suman las redes semánticas, las ontologías y los mapas conceptuales (Moreiro et al. 2002; Roger, Lavandier y Kolmayer 2001; Saadani y Bertrand-Gastaldy 2000).

La creciente demanda de este tipo de herramientas en Internet justifica la publicación en el año 2000 del estándar ISO/IEC 13250 Topic Maps que proporciona las especificaciones necesarias para crear mapas de conocimiento. Los pilares básicos de estos mapas son los conceptos (*topics*)

junto con sus relaciones (*associations*) y propiedades (*facets*). La norma define cada uno de estos componentes y establece los criterios básicos para crear relaciones y para determinar propiedades. Al estar definido para SGML o XML se facilita la navegación virtual mediante estos mapas en el entorno Web.

Aunque todas estas estructuras pueden ofrecer prestaciones similares para la recuperación de información, el análisis que sigue se centrará básicamente en las que tradicionalmente se han utilizado en los catálogos en línea. En este entorno, clasificaciones, encabezamientos de materia y en algunas ocasiones tesauros, han servido para mejorar el acceso por materias al emplearse para mostrar cómo están organizados los contenidos, hacer coincidir los términos de los usuarios con los del sistema y proporcionar términos alternativos.

1.3.4.1 Organización de la información.

Si nos atenemos al desarrollo de la actividad de búsqueda, la primera utilidad de estas estructuras es mejorar el conocimiento conceptual del usuario sobre el contenido y la organización del espacio de información con el que está interactuando. La orientación en la estructuración de contenidos facilita la navegación y la búsqueda exploratoria. Al

mismo tiempo proporciona un papel activo al usuario, aportándole una sensación de control sobre la búsqueda frente a los sistemas que la ejecutan de forma transparente (Bates 1990).

Con esta finalidad se han utilizado fundamentalmente los sistemas de clasificación y las presentaciones sistemáticas de los tesauros. La utilidad encontrada a las clasificaciones deriva de la claridad de la estructura jerárquica, mientras que la capacidad de los sistemas analítico sintéticos, aunque defendida desde el punto de vista teórico, no ha tenido aceptación en aplicaciones prácticas más allá de diseños locales y experimentales (Riesthuis 1998). Gödert (1991) defiende la superioridad de los sistemas facetados en la medida en que cada faceta puede tratarse de manera independiente en la búsqueda. No obstante, la estructura facetada puede resultar poco familiar a los usuarios, aunque como señalan Chan (2000) o Iyer (1995) esta barrera se eliminaría con una buena interfaz, con lo cual no debería ser un argumento de peso a la hora de explicar su escasa utilización.

Otra posibilidad para facilitar la navegación es fundamentarla en las relaciones entre los diferentes elementos informativos del catálogo. Un sistema pionero fue el HYPERCATalog propuesto por Hjerppe (1989), que se basaba en la exploración y la navegación como principales modos de utilización del catálogo. Para soportar la navegación el sistema tenía cinco clases de enlaces diferentes: enlaces entre registros, entre campos de diferentes registros, enlaces desde un campo hasta un registro, enlaces

desde un registro hasta un campo y campos ligados entre sí para formar un registro. Actualmente, muchos catálogos en línea accesibles vía Web facilitan algunos de estos enlaces aunque no se trate, como ya hemos señalado, de verdaderos enlaces hipertextuales.

Aparentemente, un sistema hipertextual resulta muy adecuado para facilitar las búsquedas exploratorias. Sin embargo, tampoco está exento de problemas. Por ejemplo, aunque los enlaces puedan mantener una rica red de asociaciones, sólo se puede seguir cada vez un camino, de una manera lineal. Para realizar una búsqueda completa el sistema debería ejecutar automáticamente una búsqueda paralela en las múltiples ramificaciones posibles para un mismo enlace. A esta limitación se suma el peligro de desorientación que supone la navegación hipertextual y que hace necesario que el sistema incluya algún medio para indicar al usuario los pasos seguidos hasta llegar a determinada información (Cheung 1998; Fidel et al.1999; García Marco 1994; Nelson 1992; Yan et al.1996)

1.3.4.2 Enriquecimiento de la terminología.

Los sistemas de organización del conocimiento también se han empleado en recuperación de información para enriquecer la terminología. Si se explotan

adecuadamente en entornos automatizados pueden servir para hacer coincidir los términos de los usuarios con los del sistema. Así mismo, permiten visualizar las relaciones entre los términos de manera que proporcionan puntos de acceso alternativos para ampliar, especificar o modificar la búsqueda y ayudar a descubrir nuevas materias relacionadas. Esta información puede emplearse para elaborar estrategias de ampliación de la búsqueda de forma manual o automatizada, incluyendo términos relacionados que no aparecían en el enunciado original.

La mayor parte de las soluciones propuestas en este sentido insisten en la necesidad de ampliar las relaciones de las listas de encabezamientos y traspasarlas a los registros de autoridades (Dalrymple y Younger 1991). A mediados de los ochenta Cochrane (1986) planteaba una serie de modificaciones necesarias para mejorar el uso de los encabezamientos de materia en los catálogos en línea: más relaciones de equivalencia para aumentar el vocabulario de entrada y más relaciones jerárquicas y asociativas fundamentadas en criterios lógicos. Publicaciones más recientes, como los *Principles underlying subject headings* de la IFLA (1999) o el informe del Subject Analysis Committee de la Association for Library Collections and Technical Services (ALCTS/CCS 1997), inciden en la necesidad de mejorar la estructura sindética de los encabezamientos estableciendo relaciones basadas en criterios lógicos. Concretamente, la ALCST propuso un listado con 165 posibles tipos de relaciones fundamentadas en las conexiones lingüísticas (Michel 1997). No obstante, se consideraba que sólo

algunas resultan útiles para los lenguajes de indización y éstas coinciden básicamente con las recogidas en la norma UNE 50-106-90.

Si se quiere aprovechar el potencial de las relaciones semánticas de las listas de encabezamientos de cara a la búsqueda de información en los catálogos en línea, es fundamental gestionar adecuadamente el catálogo de autoridades (Tartaglia 2003; Tillet 2000). Recientemente Tillet (2003) recordaba que también en el entorno Web el control de autoridades proporciona precisión a las búsquedas, facilita la navegación por la estructura sindética, ofrece al usuario final explicaciones sobre las variaciones que pueden parecerle inconsistencias y permite que nombres, títulos y expresiones relacionadas aparezcan juntas en las pantallas de visualización de resultados.

Sin embargo, las investigaciones realizadas hace pocos años por Cochrane (2000) y por Greenberg (1997) demostraban que la mayor parte de los sistemas no permiten que el usuario explote totalmente las relaciones que aparecen en los registros de autoridades de materias. Paulatinamente se van incorporando las relaciones de equivalencia, pero no de forma automática sino indicando al usuario cuál es el término admitido para que éste decida si emplearlo o no. En cuanto a las relaciones jerárquicas y asociativas, sólo se muestran en algunos catálogos para que el usuario conozca otros términos. En los escasos casos en los que aparecen son inconsistentes y poco manejables, además de

presentarse con una terminología poco familiar. Esta situación no parece haber cambiado sustancialmente en la actualidad.

Para incrementar la terminología también es necesario que se utilice toda la información que aparece en el registro bibliográfico y que puede expresar el contenido informativo del documento: códigos de clasificación, encabezamientos de materia y título (Cochrane 2000; Milstead 1999). Una forma de hacerlo, enriqueciendo la información de cada uno de los campos, es crear enlaces entre los índices para realizar búsquedas “cruzadas”. Por ejemplo, tanto la edición electrónica de la DDC como la de la LCC, además de incluir como puntos de acceso las rúbricas de los códigos, establecen enlaces entre estos y los encabezamientos de la LCSH (Vizine-Goetz 2001). Si esta información se incorpora a los registros de autoridad para clasificación y materias, y se establecen enlaces entre ellos, se incrementaría notablemente tanto la terminología como la información semántica a la que tienen acceso el usuario cuando realiza una búsqueda.

Otra propuesta para mejorar las posibilidades de que las búsquedas temáticas recuperen registros potencialmente relevantes es realizar enlaces desde los términos del lenguaje natural a los del vocabulario controlado que emplea el sistema. Esta solución sirve de fundamento al “tesauro enriquecido” propuesto por Bates (1986), que incluye referencias entre el vocabulario tomado del sistema de clasificación, de los títulos y de

los descriptores del documento. Uno de los sistemas pioneros que apuntaba en esta dirección fue ALEX-DOC (Hildreth 1989), un *front-end* para la recuperación en lenguaje natural. Realizaba un análisis lingüístico para equiparar los términos del usuario con los de un tesoro y elaborar así un planteamiento de búsqueda utilizando la lógica booleana.

Actualmente existen sistemas comerciales como IO Dyne que permite visualizar en diferentes pantallas tanto las relaciones que tiene en el tesoro el término de búsqueda como términos y expresiones que habitualmente coocurren con el enunciado en títulos y resúmenes. En la misma dirección, el Metadata Research Program⁴² (University of California, Berkeley) desarrolla un sistema piloto que permita vincular los términos de búsqueda del lenguaje natural con los términos de indización de vocabularios controlados utilizados en metadatos y registros bibliográficos. (Buckland et al. 1999). Este trabajo representa una confluencia de tres líneas de investigación que explotan la vinculación del lenguaje natural con tesauros (Kim y Norgard 1998; *Search support for unfamiliar metadata vocabulaires*), con encabezamientos de materia (Plaunt y Norgard 1997) y con códigos de clasificación (Larson 1991a; 1996).

⁴² Información disponible en: <http://metadata.sims.berkeley.edu/ResearchAreas/NaturalLan.html>
[Consulta 6 de abril de 2003]

1.3.4.3 Interconexión y compatibilidad.

El desarrollo de catálogos colectivos virtuales y la normalización del acceso a recursos distribuidos en Internet está poniendo de relieve la necesidad de herramientas semánticas que faciliten la interconexión de sistemas (Chan, Lin y Zeng 1999). ONE (OPAC Network in Europe) es un proyecto del Plan de Bibliotecas de la Unión Europea cuyo objetivo general es ampliar la red de información europea enlazando sistemas de catálogos nacionales de bibliotecas mediante estándares abiertos y redes de telecomunicaciones. Con el objetivo de alcanzar la interoperabilidad utilizando el protocolo Z39.50 se han considerado necesarios los perfiles de datos, los acuerdos internacionales y la cooperación para definir los registros de autoridades (Holm 1999). Sólo así se podrán ofrecer servicios de calidad que permitan al usuario: utilizar una única interfaz para acceder a sistemas remotos, ser específico en la búsqueda, visualizar los registros recuperados en diferentes formatos, visualizar los índices antes de formalizar la búsqueda y contar con asistencia por parte de los sistemas remotos concerniente a los criterios de búsqueda o al contenido de las bases de datos.

Recientemente Coyle (2000) ha realizado un estudio sobre los resultados de búsqueda en el catálogo colectivo de las bibliotecas de la University of California, Berkeley (MELVYL) empleando el protocolo Z39.50. Los resultados demuestran que los

problemas en la recuperación se deben a la falta de uniformidad de los diferentes catálogos tanto en relación a las herramientas de búsqueda como al contenido de los registros. Para que un catálogo virtual sea funcional, las bases de datos participantes deben ofrecer un conjunto uniforme de índices y de opciones de búsqueda que permitan recuperar documentos comparables de cada catálogo. Esto significa que el primer paso para la creación de un catálogo colectivo virtual es crear catálogos locales con puntos de acceso compatibles.

La compatibilidad de los vocabularios controlados ha despertado interés durante décadas como demuestra la bibliografía seleccionada por Dahlberg (1996) donde se recogen más de 500 documentos sobre el tema publicados en los últimos 40 años. Los años 70 conformaron la década más productiva y, después de un ligero declive en los 80, se convierte en un tema de actualidad debido, entre otras cuestiones, al creciente número de bases de datos accesibles vía Web. Maniez (1997) señala que pese al interés que ha despertado, el sueño de una comunicación universal entre dichos lenguajes es el paraíso perdido de los científicos de la información. Este autor defiende dos tipos de soluciones viables: la armonización de los diferentes lenguajes de información – un proceso nada fácil y además costoso – y la armonización de las fórmulas de indización a través de las tablas de concordancia – una solución más fácil que puede verse obstaculizada por discrepancias estructurales. Actualmente se considera que ésta última

es la mejor solución al problema de la incompatibilidad e integración (Nicholson y Neill 2001; MacEwan 2000; Olson 2002; Riesthuis 2001)

La idea de establecer conexiones entre ficheros de autoridades en diferentes lenguas y elaborados con distintos vocabularios controlados ha sido objeto de diversos proyectos de investigación (Chan y Zeng 2002). En una situación ideal el usuario debería poder consultar todos los catálogos en su propia lengua y recuperar documentos escritos o indizados en otros idiomas. Por tanto, en este caso el objetivo que prevalece es dar consistencia a la normalización establecida en los distintos registros, mientras que aprovechar la estructura sindética del catálogo de autoridades es un beneficio añadido (Clavel-Merrin 1999).

En el seno de la IFLA, el trabajo del IFLA UBCIM Working Group on Minimal Level Authority Records (MLAR) tiene como objetivo facilitar el intercambio determinando los elementos informativos imprescindibles en el registro de autoridad para posibilitar la consulta simultánea de diferentes catálogos de autoridades vía Z39.50. Los datos que se han considerado necesarios son: encabezamiento admitido, normas utilizadas para redactarlo, entidad responsable, lengua, nacionalidad de la entidad, referencias de véase y de véase además, y la fuente de origen. El grupo de trabajo FRANAR (Functional Requirements and Numbering for Authority Records) está

explorando además la posibilidad de establecer un número internacional para datos de autoridad (ISADN) que simplificaría el proceso de intercambio (Bourdon 2001).

A la misma iniciativa integradora responden proyectos como VIAF⁴³ (The Virtual International Authority File) de OCLC en el que participan la Library of Congress y la Deutsche Bibliothek con el objetivo de relacionar sus registros de autoridad de nombres personales. En la Unión Europea, desde mediados de los noventa está en marcha el programa CoBRA+ (Computerised Bibliographic Record Actions⁴⁴) de la DGXIII. Esta acción concertada de las agencias y bibliotecas nacionales europeas ha desarrollado los proyectos AUTHOR para autoridades de nombres de personas y entidades (Zillhardt y Bourdon 1998) y MACS (Multilingual access to subject) para encabezamientos de materia.

El trabajo del grupo del programa COBRA+ para el acceso multilingüe a la información temática (MACS)⁴⁵ es un estudio realizado entre la Swiss National Library (SNL), Deutsche Bibliothek (DDB), Bibliotheque Nationale de France (BNF) y British Library (BL) con la finalidad de establecer enlaces entre sus ficheros de autoridad (Clavel-Merrin 2003). Con este objetivo se emprendió la tarea de establecer enlaces

⁴³ <http://www.oclc.org/research/projects/viaf/index.html> [Consulta: 15 de abril de 2003]

⁴⁴ <http://portico.bl.uk/gabriel/cobra/> [Consulta: 15 de abril de 2003]

⁴⁵ "CoBRA+ Working Group on Multilingual Subject Access". Bern, March 1999 <http://portico.bl.uk/gabriel/projects/pages/cobra/finrap3.html> [Consulta: 15 de abril de 2003]

entre las tres listas de encabezamientos empleadas en las cuatro instituciones: RAMEAU (Francia y Suiza), RSWK (Alemania) y LCSH (Gran Bretaña y Suiza).

El proyecto se inició en 1997 y su primer resultado fue el desarrollo de un prototipo que proporciona acceso a 30.000 registros con encabezamientos multilingües sobre teatro y deporte. Permite que el usuario interroge en su propio idioma y recupere encabezamientos en las otras lenguas. Una de las perspectivas de aplicación de MACS es el proyecto RENARDUS, iniciado en enero de 2000. Se trata de permitir la interrogación simultánea en la Web de diferentes repertorios temáticos de recursos de Internet previamente evaluados y seleccionados. Esto significa normalizar su descripción, trabajar sobre la forma de indización y establecer medios de acceso, preferentemente multilingües.

1.3.4 La interfaz.

La situación de las interfaces para la búsqueda de información en la actualidad refleja los avances mutuamente dependientes de la investigación en Biblioteconomía y Documentación y en interacción hombre-máquina. Se han producido muchos cambios desde que se planteó por primera vez la cuestión en el *Workshop on The user interface for interactive search of bibliographic databases* (Walker 1971). A principios del siglo

XXI, las interfaces de usuario cuentan con diversos dispositivos de entrada, presentaciones gráficas en pantallas divididas en múltiples ventanas y estilos de interacción que requieren estrategias de selección y exploración más que órdenes y preguntas. (Helander, Landauer y Prabhu 1997; Joint 2001; Marchionini y Komlodi 1998; Shneiderman 1998). Los avances sugieren continuidad en esta dirección y que los diseñadores potenciarán nuevos dispositivos de entrada/salida, ubicuidad de acceso y técnicas avanzadas de visualización con el objetivo de conseguir interfaces:

- ✓ Que sean más directamente manipulables y que permitan combinar búsquedas analíticas con estrategias de exploración.

- ✓ Que presenten modelos alternativos y adaptables para satisfacer las necesidades de diferentes tipos de usuarios.

- ✓ Que permitan contextualizar la búsqueda de información dentro de un proceso de trabajo más amplio.

Se han propuesto múltiples criterios para mejorar la amigabilidad: los sistemas deberían ser fáciles de entender, fáciles de aprender, tolerantes con los errores, flexibles, adaptables, potentes, eficientes, baratos, compatibles, inteligentes, seguros y fáciles de usar (Computer Science and Telecommunications Board 1997). Aunque

algunos son particularmente aplicables al diseño de catálogos en línea, muchas mejoras han sido sólo superficiales porque la estructura de la base de datos, la de los índices y las técnicas de recuperación, impiden transformaciones más profundas (Abadal 2002; Belkin et al. 1991; Espelt 1998, Ortiz-Repiso y Moscoso 1999). De las múltiples líneas de investigación que se han abierto en relación con este tema, aquí se examinan las tres que tocan más de cerca al manejo de la búsqueda temática en los OPAC: la visualización de los índices de materia, los sistemas expertos y las presentaciones gráficas.

1.3.5.1 Visualización de los índices.

Los resultados de los primeros trabajos realizados sobre el acceso por materias ponían de manifiesto que incluso para búsquedas muy específicas puede ser necesario que el usuario visualice cientos de encabezamientos con el objetivo de obtener una visión completa sobre un tema. Al mismo tiempo, se ha demostrado que la organización de los índices de materia no es intuitiva para los usuarios ni tiene porqué estar necesariamente de acuerdo con la forma en que piensan encontrar la información. Frecuentemente no entienden los criterios de ordenación, ni la función de los encabezamientos y subencabezamientos, ni las referencias de la estructura sindética (Drabenstott, Simcox y Fenton 1999). Además, para diseñar la visualización de los encabezamientos en la

interfaz del OPAC, es importante recordar que nunca se debe asumir que los usuarios conocen el encabezamiento que necesitan con anterioridad a realizar la búsqueda. Las recomendaciones para mejorar la navegación en los índices de encabezamientos precoordinados han sido un tema habitual en la bibliografía profesional durante más de dos décadas. La finalidad de todas ellas ha sido reducir el esfuerzo cognitivo que los usuarios tienen que hacer cuando realizan una exploración del sistema.

Cochrane (1986) y Massicote (1986) fueron pioneras en proponer la “contracción”⁴⁶ como un medio para hacer que los índices de encabezamientos de materia fueran más fáciles de visualizar. Sugerían que los encabezamientos de materia se presentaran indicando simplemente la entrada principal y la existencia de subdivisiones de materia, lugar, cronológicas o de forma, de manera que el usuario decidiera qué ver. Basándose en parte en este trabajo Allen (1993), Drabenstott y Vizine-Goetz (1994) o MacGarry y Svenonius (1991) han propuesto soluciones similares para aquellas búsquedas que coinciden parcialmente con el índice y recuperan muchos encabezamientos. Se insiste, además, en que para ayudar a la exploración es importante fomentar una presentación estructurada, que refleje las relaciones sindéticas de la lista (jerarquía, asociación, equivalencia) en detrimento de la presentación alfabética que no resulta significativa la mayor parte de las veces.

A nivel institucional, en 1987 el Subcommittee on the Display of Subject Headings in Subject Indexes in Online Public Access Catalogs de la American Library Association (ALA) se encargó de estudiar los problemas de la presentación de los encabezamientos de materia en los catálogos en línea. El resultado fue la publicación de una guía (*Headings for tomorrow* 1992) que recogía las principales recomendaciones y que continúa siendo una obra de referencia en esta área.

Años más tarde, en 1997, la IFLA formó un grupo de trabajo - Task Force on Guidelines for OPAC's Displays - con el objetivo de establecer recomendaciones para la visualización en los catálogos en línea, interfaces gráficas y productos multimedia. El borrador provisional que se publicó en 1998 sigue siendo objeto de discusión actualmente (*Guidelines for OPAC displays* 1998; Yee 1999). En relación con la presentación de los encabezamientos de materia recoge las propuestas de los trabajos de la ALA y de las investigaciones anteriores. En síntesis las recomendaciones sobre la presentación de los encabezamientos de materia son las siguientes:

- ✓ El sistema debería mostrar al usuario los índices de los campos de búsqueda siempre que como resultado de la misma se recupere más de un registro.

⁴⁶ Se ha traducido el término inglés *compression* como contracción por considerarse que esta palabra tiene un significado más ajustado a su sentido contextual y es de uso más habitual que la traducción literal (compresión).

- ✓ La visualización de los índices debería responder a una ordenación estructurada en lugar de a la estrictamente alfabética, que no resulta significativa. Para implementarla se utilizan como primer criterio los de ordenación los indicadores de subcampo del formato MARC (“elementos de ordenación”). De esta manera cada encabezamiento se subdividiría primero por los subencabezamientos de materia, después por los de forma, lugar o cronológico, que posteriormente se ordenarán alfabéticamente.

Ordenación alfabética	Ordenación estructurada
Medio ambiente	Medio ambiente
Medio ambiente–África	Medio ambiente–Conservación
Medio ambiente–Almería	Medio ambiente–Cooperación internacional
Medio ambiente–Atlas	Medio ambiente–Influencia humana–Galicia
Medio ambiente–Bibliografías	Medio ambiente–Protección
Medio ambiente–Conservación	Medio ambiente–Atlas
Medio ambiente–Cooperación internacional	Medio ambiente–Bibliografías
Medio ambiente–Diccionarios	Medio ambiente–Diccionarios
Medio ambiente–Influencia humana–Galicia	Medio ambiente–África
Medio ambiente–Protección	Medio ambiente–Almería
Medio ambiente–Sevilla(Provincia)	Medio ambiente–Sevilla(Provincia)

Fig. 1.12 Visualización de índices. Ordenación de subcampos

- ✓ Cada uno de estos elementos se considera independiente para la ordenación alfabética. Además, dentro del encabezamiento principal (\$a) se reconocen los signos de puntuación como elementos diferenciadores.

Ordenación alfabética	Ordenación estructurada
Poder	Poder
Poder ejecutivo	Poder-Historia
Poder (Filosofía)	Poder-Portugal
Poder-Historia	Poder (Filosofía)
Poder judicial	Poder (Psicología)
Poder legislativo	Poder ejecutivo
Poder-Portugal	Poder judicial
Poder (Psicología)	Poder legislativo
Poder real	Poder real

Fig. 1.13 Visualización de índices. Ordenación estructurada

- ✓ La ordenación de los subencabezamientos de historia debería ser cronológica, no alfabética.

Ordenación alfabética	Ordenación estructurada
España-Historia-Alfonso XII, 1874-1885	España-Historia-Edad media, 476-1942
España-Historia-Carlos I, 1516-1556	España-Historia-Reconquista, 718-1492
España-Historia-Comunidades de Castilla, 1520-1521	España-Historia-Comunidades de Castilla, 1520-1521
España-Historia-Edad media, 476-1942	España-Historia-Carlos I, 1516-1556
España-Historia-Primera República, 1873-1874	España-Historia-Primera República, 1873-1874
España-Historia-Reconquista, 718-1492	España-Historia-Alfonso XII, 1874-1885

Fig. 1.14 Visualización de índices. Orden cronológico

- ✓ Se debería contraer los encabezamientos cuando los resultados de la búsqueda sean demasiado amplios.

Visualización con contracción	
Medio ambiente	Si se selecciona la línea 1 se visualizaría
1 Subdividido por áreas geográficas (Continentes, países, comunidades autónomas, etc.)	Medio ambiente–África Medio ambiente–Almería Medio ambiente–Sevilla(Provincia)
2 Subdividido por formas de presentación (Diccionarios, enciclopedias, anuarios, etc.)	<i>Si se selecciona la línea 2</i>
3 Subdividido por periodos cronológicos	
Medio ambiente–Conservación	Medio ambiente–Atlas
Medio ambiente–Cooperación internacional	Medio ambiente–Bibliografías
Medio ambiente–Protección	Medio ambiente–Diccionarios

Fig. 1.15 Visualización de índices. Contracción

- ✓ Se deberían mostrar las relaciones de los términos en el índice alfabético de manera que se ponga de manifiesto la estructura sindética del catálogo.

Electrones

Términos genéricos		Términos relacionados	
1	Átomos	14	Iones
2	Leptones	15	Positrones
3	Materia-Estructura	16	Rayos catódicos
4	Partículas(Física nuclear)		
Términos específicos			
5	Cañones electrónicos		
6	Carga espacial		
7	Configuración electrónica		
8	Efecto Auger		
9	Electrones-Distribución		
10	Fotoelectrones		
11	Mesones(Partículas)		
12	Orbitales atómicos		
13	Rayos beta		

Fig. 1.16 Visualización de índices. Relaciones semánticas

Sin embargo, las recomendaciones realizadas a lo largo de estos años parecen haber tenido escasa repercusión en el diseño de los actuales OPAC's (Greenberg 1997). Recientemente, Wool (2000) ha analizado las presentaciones de quince catálogos accesibles vía Web que representan seis de los grandes distribuidores de sistemas de gestión bibliotecaria de Estados Unidos: Ameritech, DRA, Endeavor, Innovative Interfaces/III, Library Corporation/TLC, Sirsi. La mayor parte de ellos utilizan la ordenación alfabética sin tener en cuenta los tipos de subdivisión o los calificadores.

Los sistemas que emplean índices de encabezamientos permutados - Innovative y Sirsi - presentan los resultados en una única secuencia alfabética que resulta muy confusa para los usuarios. Además, actualmente muchos catálogos en línea ejecutan una búsqueda por palabras clave en el campo materias y presentan como resultado directamente los registros recuperados en formato abreviado⁴⁷. Evidentemente, estos dos últimos tipos de búsqueda permiten mitigar algunos de los problemas que plantea la sintaxis de los encabezamientos de materia. Sin embargo, su implementación no debería significar una renuncia a la visualización de los índices de materia con una presentación clara y estructurada

⁴⁷ Sólo uno de ellos (Ameritech) presenta una lista de encabezamientos como repuesta a la búsqueda por palabras clave en el encabezamiento, desgraciadamente en un orden no reconocible (Wool 2000).

1.3.5.2 Sistemas expertos.

Además de proporcionar una visualización estructurada de los índices, el catálogo en línea debería facilitar más mensajes de ayuda a los usuarios sobre cómo proceder si tienen alguna dificultad. Los mensajes deberían decirle al usuario qué hacer, cómo hacerlo y cómo puede mejorar los resultados (Hildreth 1987). Cuando no se recuperan registros o se recuperan pocos, el sistema puede sugerir que se acorte la frase de búsqueda, que se sustituyan los términos iniciales por sinónimos o términos más generales o que se vuelva a iniciar la búsqueda utilizando un método diferente. Cuando se recuperan demasiados registros, puede solicitar al usuario que añada términos de búsqueda adicionales o que incorpore nuevos criterios que reduzcan la búsqueda. También puede proponer al usuario que intente una búsqueda diferente y sugerirle opciones que desconocen (Vickery y Vickery 1993).

Paradójicamente, el gran problema de los mensajes de ayuda puede ser que el usuario se vea desbordado por la información que proporcionan y decidan no utilizar el OPAC (Borgman 2000). Por tanto, sólo deberían proporcionar una cantidad de ayuda limitada. Una posible solución para este problema es utilizar un sistema experto en búsqueda bibliográfica que reúna el conocimiento y las habilidades de expertos, de forma que sea capaz de mimetizar su comportamiento y tomar decisiones por el usuario (Efthimiadis

1990; Chen y Dhar 1991). Además de reglas y datos, la base de conocimiento de un sistema experto intermediario debería contener estrategias que permitan aclarar el tema de búsqueda, estrategias para la recuperación de información y reglas para seleccionar las estrategias.

La mayoría de los sistemas expertos intermediarios que se han desarrollado están pensados para sistemas de recuperación de información que manejan resúmenes o texto completo: THOMAS (Jamieson y Oddy 1979), RABBIT (Williams 1984), RUBRIC (Tong, et al. 1985), PLEXUS (Vickery, et al. 1987), I3R (Croft y Thompson 1987). En diferentes ocasiones se ha planteado que los sistemas expertos sólo son útiles para resolver problemas en dominios muy específicos (Savage-Knepshild y Belkin 1999). Autores como Salton (1986) o Brooks (1987) incluso han manifestado dudas sobre que se pudiera desarrollar un sistema experto para recuperar información. El fundamento de su opinión radica en el hecho de que la búsqueda de información implica tareas muy heterogéneas y en que habitualmente son utilizados por usuarios con características muy diferentes.

A primera vista, esta podría no parecer una buena solución para un catálogo en línea debido a la gran cantidad de conocimiento temático que contiene y al público al que se dirige. Sin embargo, centrándose en las estrategias de búsqueda y en las reglas que pueden utilizarse para realizarla de forma eficaz, Khoo y Poo (1994) consideran que un

sistema experto debería resultar una solución mejor para las búsquedas en OPAC's que otros tipos de interfaces. Entre los requisitos que debería seguir el diseño del sistema experto incluyen:

- ✓ Incorporar un repertorio de estrategias derivadas de la experiencia de los profesionales.
- ✓ No ejecutar las estrategias en una secuencia fija, sino incluir reglas para seleccionarlas en función de las características de la búsqueda.
- ✓ Explicar la estrategia de búsqueda que ha utilizado y por qué la ha seleccionado.
- ✓ Incorporar una base de conocimiento modular, de forma que se puedan añadir o modificar estrategias con facilidad.
- ✓ Registrar su propio funcionamiento y utilizarlo en búsquedas posteriores.

1.3.5.3 Interfaces gráficas.

Otra opción que se ha planteado para facilitar la interacción es la utilización de interfaces gráficas que representen de manera metafórica el espacio de información en el que se mueve el usuario. La metáfora que se utiliza es una representación que permite interactuar de forma sencilla y por analogía con un sistema mucho más complejo de manera totalmente transparente al usuario, pero que no deja de ser

representación necesariamente incompleta y sesgada de dicho sistema (Dieberger 1997, Eberts 1994).

Los VIRI's (*Visual Information Retrieval Interfaces*) suponen una vía alternativa porque favorecen la utilización de las capacidades cognitivas de reconocimiento espacial frente a los sistemas tradicionales que emplean las conceptuales al trabajar con el lenguaje y con la lógica. Se trata de presentar la información de manera que sea posible asimilarla por el sistema perceptual humano en lugar de recaer sobre el sistema cognitivo. Aunque no está muy claro todavía en qué medida pueden ayudar a mejorar las búsquedas por materias, sí se ha demostrado que reducen el tiempo que el usuario emplea cuando realiza búsquedas exploratorias (Allen 1998, Beheshti 1992)

Aunque la mayoría utilizan sistemas postcoordinados más que precoordinados para crear las categorías, constituyen una línea de investigación que puede permitir presentar al usuario sistemas precoordinados de forma más manejable. Por ejemplo, Borgman (1995) y Beheshti (1992) proponen utilizar los códigos de clasificación para crear una metáfora que simule la colocación de los libros en las estanterías, de manera que el usuario pueda realizar un “paseo virtual” por la biblioteca. Las investigaciones promovidas por OCLC en el marco del proyecto *Information visualization*⁴⁸ utilizan

⁴⁸ <http://research.oclc.org/Normore/public/OverView/index.html> [Consulta: 24 de abril de 2003]

tanto términos extraídos del lenguaje natural (WordView3D⁴⁹) como de la DDC (Barcelona⁵⁰ y FullView⁵¹) para representar gráficamente los resultados de la búsqueda.

La mayor parte de las investigaciones se orientan a elaborar las representaciones gráficas de manera automatizada en sistemas de recuperación que trabajan con texto completo o con resúmenes de los documentos⁵². Moya Anegón y Herrero Solana (1998) las dividen en dos grupos en función de las dimensiones que utilizan para crear la metáfora del espacio de información.

Entre las representaciones en dos dimensiones destacan los SOM (Self-Organizing Map) de Kohonen (1997) que clasifican automáticamente la información de entrada, creando mapas topográficos del contenido de la base de datos. Uno de los prototipos que utilizan esta aplicación es WEBSOM⁵³, un proyecto dirigido por el propio Kohonen en la Universidad Tecnológica de Helsinki y que clasifica grandes volúmenes de información procedentes de un grupo de noticias Usenet.

El sistema tiene una interfaz muy sencilla de manejar en la que las áreas oscuras indican mayor densidad de información que las claras, y los temas aparecen resumidos

⁴⁹ <http://research.oclc.org/Normore/public/OverView/WordView3D/wv3.asp> [Consulta: 24 de abril de 2003]

⁵⁰ <http://research.oclc.org/Normore/public/OverView/Barcelona/barcelona.asp> [Consulta: 24 de abril de 2003]

⁵¹ <http://research.oclc.org/Normore/public/OverView/FullView/fv.asp> [Consulta: 24 de abril de 2003]

en tres letras. Cuando se selecciona una determinada zona se accede primero a un mapa ampliado de la misma y luego a una lista de documentos relacionados temáticamente (Honkela, et al. 1996; Kaski, et al.1996; Lagus, et al. 1996).

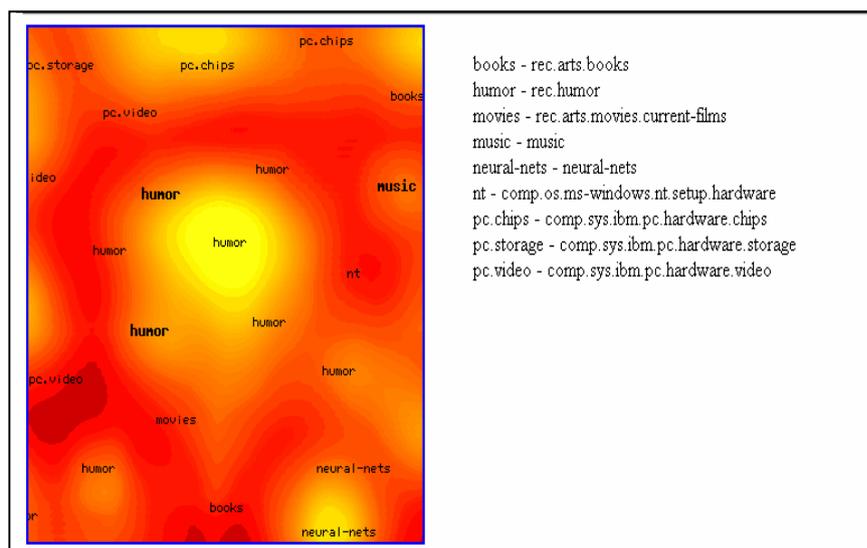


Fig 1.17 Interfaces gráficas. Proyecto WEBSOM

Los desarrollos de metáforas en 3D, se pueden dividir en dos grandes grupos: las basadas en nodos espaciales y las basadas en mapas estructurados. Las primeras son muy utilizadas para navegar entornos hipertextuales, a través de racimos de nodos unidos mediante líneas de conexión, mientras que las segundas son más indicadas para representar relaciones estructurales más complejas como el contenido documental de

⁵² Se puede encontrar una síntesis de los proyectos y prototipos desarrollados en: The Big Picture(sm) <http://www.public.iastate.edu/~CYBERSTACKS/BigPic.htm> [Consulta: 24 de abril de 2003]

⁵³ <http://websom.hut.fi/websom>. [Consulta: 24 de abril de 2003]

una base de datos. Un claro ejemplo del primer grupo lo constituye el modelo Narcissus, desarrollado en la Universidad de Birmingham, y que tiene como objetivo la organización tridimensional de la información de la Web (Hendley 1995). Para ello forma una suerte de galaxias y constelaciones de esferas de distinto tamaño unidas por líneas. Las esferas representan las páginas y las líneas los enlaces. Basándose en este modelo se ha desarrollado una aplicación especial denominada HyperSpace⁵⁴ (Fig. 1.?), que también esta dirigida a la gestión de información en Internet (Wood 1995).

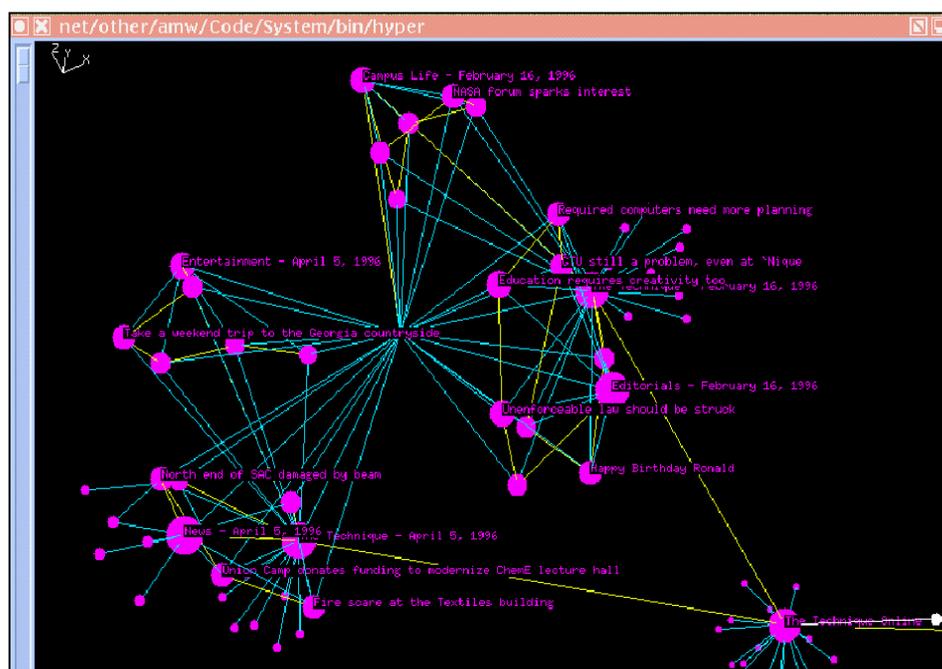


Fig. 1.18 Interfaces gráficas. HyperSpace

⁵⁴ <http://www.cs.bham.ac.uk/~nsd/> [Consulta: 24 de abril de 2003]

Un ejemplo del segundo grupo, los mapas estructurales, lo encontramos en un prototipo denominado LyberWorld (Hemmje 1995), que consiste en una interfaz de usuario desarrollado para un sistema probabilístico de recuperación en texto completo (INQUERY). LyberWorld⁵⁵ presenta dos metáforas: las esferas de relevancia, dirigidas a la búsqueda (figura 1.19) y los conos de navegación, utilizados para la exploración de la base de datos (figura 1.20).

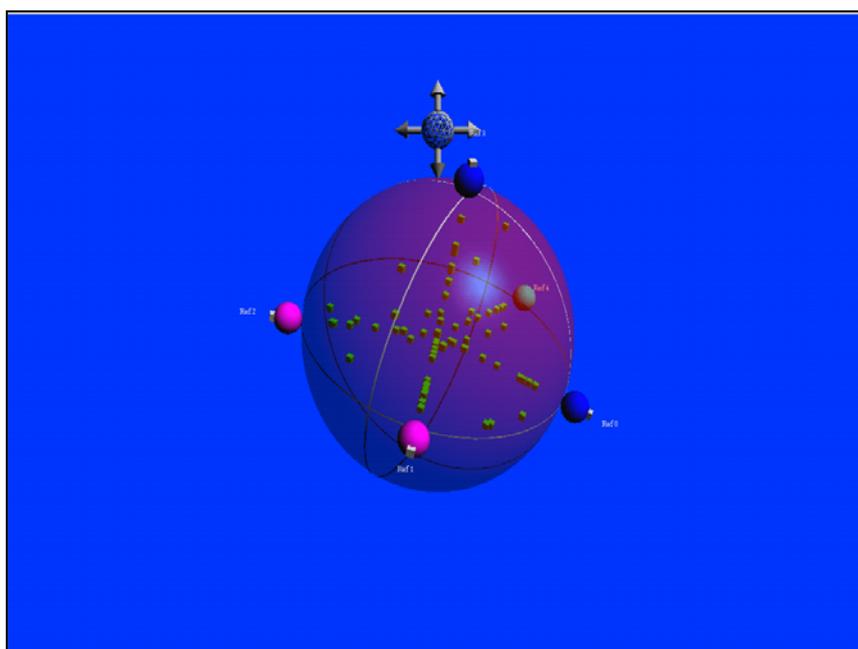


Fig. 1.19 Interfaces gráficas. LiberWorld: esferas de relevancia

⁵⁵ <http://www-cui.darmstadt.gmd.de/visit/Activities/Lyberworld/demos.html> [Consulta: 24 de abril de 2003]

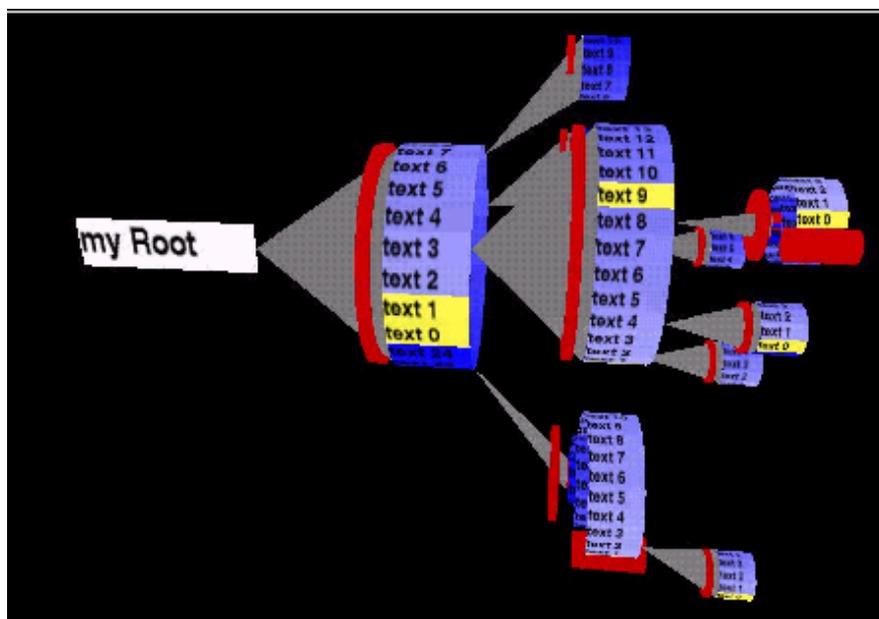


Fig. 1.20 Interfaces gráficas. LiberWorld: conos de navegación

1.4 Líneas de investigación.

En los últimos años se han publicado numerosas investigaciones en las que se discute cuál es el modelo teórico más adecuado para la Documentación: positivista, cognitivo, hermenéutico, sociológico, naturalista, etc. (Fernández Molina y Moya Anegón 2002; Hjørland 2000; Pettigrew, Fidel y Bruce 2001). Esta mezcla de paradigmas se debe tanto la carencia de teorías propias, que definan los conceptos básicos y sus relaciones, como a las dudas sobre la orientación metodológica más

adecuada (Dick 1999; Vakkari y Kuokkanen 1997). Como consecuencia, los distintos puntos de vista parecen “darse codazos” para convertirse en el único enfoque válido, lo que ha provocado como consecuencia negativa la sustitución acrítica de unos por otros (Zwadlo 1997).

Evidentemente, esta situación se reproduce en el ámbito específico de la investigación sobre recuperación de información. Como señalaban Robertson y Hancock-Beaulieu (1992), las diferencias fundamentales en esta área se producen por la cuestión de dónde situar los límites del “sistema” de búsqueda. Las diferencias a la hora de emplazarlos han servido para distinguir dos líneas de investigación fundamentales en esta área: un enfoque basado en el sistema - tradicional o algorítmico- y un enfoque centrado en el usuario o cognitivo (Ellis 1992).

Para el enfoque tradicional el “sistema” es el mecanismo que permite almacenar y recuperar la información, la estrategia de búsqueda es el input y los documentos recuperados el output. Esta delimitación contrasta con la situación que se produce en el mundo real, donde las necesidades de información surgen en contextos concretos y las personas están implicadas en el proceso de recuperación. La perspectiva “centrada en el usuario” amplía los límites del “sistema” con la incorporación del componente humano y del entorno en el que se produce la búsqueda. El cambio de perspectiva se relaciona con diferentes factores de carácter pragmático y teórico:

- Por una parte, al popularizarse el uso de las redes de información, la recuperación se ha desplazado desde el terreno de los especialistas al de los usuarios finales. Esta nueva población de usuarios, cada vez más numerosa y heterogénea, reclama mecanismos que se adapten sus habilidades y necesidades: herramientas interactivas y amigables que proporcionen acceso a información multimedia facilitando diferentes tipos de búsqueda (equiparación, exploración, navegación). Para diseñar estos sistemas es necesario saber qué demandan los usuarios y cómo se comportan durante la búsqueda. (MIRA...2000)
- Paralelamente, se ha producido un cambio en el concepto de la naturaleza de la interacción. Ya no se considera que el *input* sea una estrategia de búsqueda sino un usuario con una necesidad de información, con un estado anómalo de conocimiento (Belkin, Oddy y Brooks 1982). En el extremo opuesto, el *output* no es un conjunto de documentos potencialmente relevantes sino un usuario cuyo estado de conocimiento se ha modificado durante la interacción.

El interés actual por la investigación sobre la interacción usuario/sistema parece favorecer una aproximación de las dos perspectivas. Sin embargo, se trata de un acercamiento complejo, porque supone la confluencia de enfoques muy diferentes tanto

en sus fundamentos conceptuales como en sus planteamientos metodológicos (Harter y Hert 1997; Olvera Lobo 1999).

Para la perspectiva tradicional, el objetivo primordial de la investigación es mejorar las técnicas de recuperación y los métodos de representación de la información, de manera que se facilite la equiparación entre búsquedas y documentos. Los trabajos de investigación simulan un entorno de búsqueda artificial, controlado, formado por cuatro componentes: un conjunto de documentos, un sistema de almacenamiento y recuperación, un conjunto de temas que se concretan en enunciados de búsqueda y un conjunto de juicios sobre la relevancia de los documentos para esos enunciados.

La eficacia de los sistemas se evalúa en función de las medidas de exhaustividad y precisión basadas en la relevancia temática que se establecieron en Cranfield (Cleverdon 1967). Esta evaluación se realiza en un entorno en el que se controlan las variables, para asegurar que los efectos constatados en la investigación se deben a cambios en alguno de los parámetros del sistema. Aunque la metodología que se emplea está bien consolidada, ha sido criticada al menos en dos aspectos: su limitación en tamaño y escala cuando se compara con sistemas comerciales, y la carencia que supone el hecho de no tener en cuenta la contribución humana al proceso de búsqueda.

La crítica sobre tamaño y escala se ha abordado en las Text REtrieval Conferences (TREC) empleando colecciones experimentales que contienen varios gigabytes de datos al tiempo que se utilizan temas de búsqueda que intentan reproducir situaciones reales (Harman 1995; Over 2001). Aunque no fue su objetivo original, estos experimentos también han afrontado la segunda crítica, la eliminación del componente humano, introduciendo una línea de trabajo “interactiva” (*TREC Interactive Track*). En esta modalidad, los participantes comprueban la eficacia de diferentes técnicas para mejorar la interacción usuario/sistema siguiendo las directrices generales de las conferencias, de forma que los resultados sean comparables (Beaulieu, Robertson & Ratmussen 1996).

Del otro lado, la perspectiva “centrada en el usuario” se inicia con los trabajos de Zweizig y Dervin (1977) para quienes la búsqueda y utilización de la información debe analizarse dentro del contexto sociocultural en el que se produce y se resuelve. Básicamente, se trata de conocer las características contextuales e individuales que influyen en la formalización de las necesidades de información, en el desarrollo del proceso de búsqueda y en la valoración de los resultados obtenidos. Bajo este prisma, el rendimiento de los sistemas se evalúa utilizando como criterio la “relevancia basada en la situación”, que considera no sólo la coincidencia temática entre búsqueda y documento sino la utilidad que estos tienen para el usuario (Harter 1992, Peis y Fernández Molina 1994, Scamber 1994).

En realidad, la etiqueta “centrada en el usuario” ha servido para agrupar un conjunto bastante diverso de enfoques que tienen en común la presencia del “factor humano” y la oposición a la perspectiva tradicional (Allen 1996, Dalrymple 2001, Nahl 1998, Savolainen 1993). Por ejemplo, Sugar (1995) señala dos orientaciones: una cognitiva, que estudia cómo se procesa la información y otra holística que tiene en cuenta tanto aspectos afectivos y psicomotores como contextuales. Sin embargo, Hert (1997) plantea tres enfoques en función del objeto principal de estudio y de la metodología empleada: uno que se centra en la relevancia contextual; otro que analiza la influencia de las características individuales en el comportamiento de búsqueda – que incluye un enfoque cognitivo - ; y el que examina los cambios que se producen en los objetivos y en las estrategias de los usuarios desde una perspectiva “naturalista”. Por su parte, Fidel (2000) diferencia dos generaciones dentro de los estudios que se reúnen bajo esta denominación: la primera, analiza la influencia que tienen de las características de los usuarios sobre la interacción; y la segunda, trata de identificar patrones de comportamiento en este proceso.

La confusión conceptual y terminológica que ha provocado esta variedad de puntos de vista tiende a simplificarse actualmente reconociendo dos corrientes epistemológicas fundamentales (Fernández Molina y Moya Anegón 2002; Pettigrew, Fidel y Bruce, 2001; Sonnenwald y Iivonen 1999):

- a) La sociológica, que se centra en el análisis de las influencias “contextuales” - políticas, sociales, institucionales o culturales – sobre el comportamiento de los usuarios de sistemas de recuperación de información.

- b) La cognitiva, que aborda el estudio de los aspectos emocionales y cognitivos de esta conducta que son independientes del contexto en el que se realiza la búsqueda.

Para este último enfoque, que surge a partir de los trabajos de De Mey (1977), entender la búsqueda de información significa comprender cómo las categorías conceptuales de los usuarios – su modelo del mundo - afectan a la interacción y se modifican durante la misma. Sus planteamientos han tenido tanta importancia en la investigación sobre recuperación de información que muchos autores consideran que es el “paradigma cognitivo” el que sirve de verdadero contrapunto a la perspectiva tradicional (Ellis 1992; Belkin 1990; Allen 1991; Ingwersen 1992, 1996).

En cuanto a los trabajos empíricos, esta perspectiva también abarca un heterogéneo conjunto de investigaciones cuyo denominador común es el análisis de situaciones de búsqueda en entornos “reales”. Para alcanzar este objetivo se ha incrementado el uso de metodologías cualitativas con las que se intenta explicar el comportamiento humano

(Fidel 1993). Sin embargo, y a pesar del interés que ha suscitado este tipo de trabajos en la última década, todavía no se ha consolidado un grupo de medidas normalizadas que permitan comparar los resultados de la investigación.

La complejidad de esta situación se pone de manifiesto en trabajos como los de Vickery (1973), Fidel y Soergel (1983) o Yuan y Meadow (1999) que recogen más de un centenar de variables sistematizándolas en categorías que abarcan: contexto, características del usuario, tutoriales, tema de búsqueda, base de datos, especialista en recuperación de información, sistema de búsqueda, características del proceso y valoración de los resultados. En otros casos la sistematización afecta sólo a alguna de estas categorías: por ejemplo, Su (1992) recopiló las variables relacionadas con el rendimiento de los sistemas de recuperación de información, mientras que Borgman (1989) lo hizo con las vinculadas a las características individuales de los usuarios.

Siguiendo la línea de estos trabajos, se ha realizado una revisión bibliográfica con objeto de saber qué metodologías y qué variables se han empleado en las investigaciones sobre búsqueda de información. Esencialmente se han seleccionado investigaciones originales sobre el comportamiento de búsqueda de usuarios de catálogos en línea. Sin embargo, se han incluido algunos trabajos en los que se analizaban las búsquedas en la WWW o en bases de datos en CD-ROM y que se

pueden aplicar al estudio de las nuevas características de los OPAC's (enlaces hipertextuales, visualización de registros, técnicas de recuperación, etc.).

Se han examinado trabajos publicados básicamente en la década de los 90, en especial en los últimos 5 años, para evitar solapamientos con otras revisiones bibliográficas sobre el tema como las de Ríos García (1991), Seymour (1991), Peters, Kurth y Kaske (1993), O'Brien (1994), Large y Beheshti (1997) o el número especial del *Journal of the American Society for Information Science* (1996). Los resultados se han organizado en dos grandes apartados: uno para las técnicas de recogida de datos y los métodos de análisis, y otro para las variables empleadas.

1.4.1 Técnicas de recogida de datos y métodos de análisis.

El primer rasgo que se puede destacar es la falta de consistencia que se constata también a la hora de delimitar y denominar métodos de investigación, técnicas de recogida de datos y métodos de análisis. Las diferencias conceptuales y terminológicas han provocado la existencia de diferentes clasificaciones, entre las que Delgado López-Cozar (2000a) destaca el esquema de Järvelin y Vakkari (1990). Sin embargo, pese a su gran utilidad, las categorías de este modelo son producto de una mezcla de criterios porque está pensado para clasificar trabajos de investigación y no metodologías

específicas. Como resultado, algunas de las “metodologías” son meras técnicas de recogida de datos, otras métodos de análisis, otras diseños de investigación y otras una combinación de varios factores.

Para evitar la confusión que puede originar esta categorización, se eligió una división más pragmática, en la línea de las utilizadas por MacClure (1994), Westbrook (1994) o Borrego Huerta (2001). En ellas se establece una división entre técnicas de recogida de datos y métodos de análisis que está relacionada con el desarrollo lógico del proceso de investigación:

- En el grupo de técnicas de recogida de datos se incluyen aquellos sistemas utilizados para registrar la información relativa a cualquier persona, sistema o proceso objeto de estudio.
- En el apartado métodos de análisis se reseñan los procedimientos empleados para analizar la información original con el objetivo de derivar resultados que permitan explicar el fenómeno estudiado.

El siguiente cuadro sintetiza el resultado de la revisión de los trabajos de investigación.

Técnica de recogida de datos	Método de análisis de datos
<ul style="list-style-type: none">• Cuestionarios• Registro de las transacciones• Grabación de los resultados• Grabación audio de la interacción durante la búsqueda (<i>Protocolos verbales</i>)• Entrevistas con los usuarios• Grupos de discusión• Observación directa	<ul style="list-style-type: none">• Métodos estadísticos<ul style="list-style-type: none">Métodos descriptivosMétodos inferenciales• Métodos cualitativos<ul style="list-style-type: none">Análisis de contenidoTeoría fundamentada (<i>Grounded theory</i>)Descripción densa

Fig. 1.21 Técnicas de recogida de datos y métodos de análisis

La grabación de las transacciones se ha convertido en la técnica de recogida de datos más utilizada en los últimos años para estudiar el comportamiento de los usuarios de sistemas de recuperación de información (Jones 1997). Esta técnica, que se ha aplicado al estudio de los catálogos en línea desde mediados de los ochenta (Frías y Martín Rodríguez 1999; Peters 1997), no es obtrusiva y proporciona una información detallada y abundante sobre la interacción usuario-sistema que permite construir modelos cuantitativos y ayudar a hacer interpretaciones cualitativas (Borgman, Hirsh y Hiller 1996). Se ha empleado para analizar características del proceso de búsqueda como puntos de acceso utilizados, errores cometidos y número de registros recuperados (Blecic et al. 1998; Larson 1991b), tiempo empleado en la sesión, pantallas visualizadas y secuencias de búsqueda (Cooper 2001; Ray y Lang 1997; Wallance 1993), términos y operadores lógicos utilizados (Drabenstott y Vizine-Goetz 1994; Ensor 1992; Hildreth 1997; Jansen, Spink y Saracevic 2000).

Sin embargo, la riqueza de la información que aporta está limitada porque no proporciona datos sobre las características de los usuarios o el objetivo de la búsqueda, no permite saber si se busca por un ejemplar conocido o por materias, es difícil identificar dónde empieza o termina una búsqueda y sólo informa parcialmente del comportamiento de búsqueda, ya que no permite explicar el porqué de las acciones (Hancock-Beaulieu, Robertson y Neilson 1991). Estas razones justifican que habitualmente se utilice junto con otras técnicas de recogida de información de corte cualitativo: cuestionarios y observación directa (Connaway, Budd y Kochtanek 1995), cuestionarios (Hancock-Beaulieu et al. 1995) o con protocolos verbales y entrevistas estructuradas (Connell 1995).

Otro medio para obtener información de los usuarios es que completen un cuestionario en el que es posible recoger tanto información demográfica como información relacionada con la experiencia (Connaway 1995; Slone 2000), la idea que tienen del funcionamiento de sistema (Hildreth 1997) o la satisfacción con los resultados (Spink y Saracevic 1997). Una opción para distribuir los cuestionarios, que facilita la tarea del investigador y da más libertad a los usuarios, es ponerlos en línea para que se cumplimenten antes, durante o después de la búsqueda (Drabenstott 1996; Ensor 1992; Ferl y Millsap 1996; Hancock-Beaulieu et al. 1995; Snelson 1993).

Un sistema para conseguir información detallada acerca de la opinión de los usuarios sobre la búsqueda es entrevistarlos después de haberla realizado, aunque en algunos casos también se realiza una entrevista previa (Beheshi, Large y Bialek, 1996, Irgens 2000). La entrevista sirve para complementar la información de las transacciones porque permite conocer las actitudes, las opiniones de los usuarios y sus percepciones de los éxitos o los fallos cuando utilizan los sistemas de información (Hert 1996; Borgman 1995). Una alternativa a la entrevista individual es el grupo de discusión en el que las percepciones y actitudes de los miembros se ponen de manifiesto mediante la interacción de las personas que lo integran (Connaway 1997).

La observación directa es una técnica de recogida de datos que puede resultar más objetiva que el cuestionario y la entrevista porque mientras en las anteriores los usuarios expresan sus opiniones, aquí es el investigador el que registra el desarrollo de una actividad. Sin embargo, tiene el inconveniente de que la presencia del investigador puede alterar el desarrollo “natural” del fenómeno observado y en este caso puede modificar el comportamiento de los usuarios. Una alternativa es la observación encubierta, que ha permitido, por ejemplo, establecer el principio y el final de las búsquedas registradas en las transacciones (Wiberley et al. 1995), o la grabación en vídeo de la sesión de búsqueda (Kiestra, Stokmans y Kamphuis 1994).

Pese a las críticas que ha suscitado, esta técnica proporciona una información muy valiosa sobre el proceso de búsqueda y, especialmente, sobre las reacciones y emociones de los usuarios durante el mismo (Fidel 1999; Slone 2000). Una información similar puede obtenerse mediante los protocolos verbales en los que se registran los pensamientos que el usuario expresa en voz alta mientras realiza la búsqueda. En este caso se trata de su opinión sobre el proceso, la relevancia de los registros recuperados, la modificación de la estrategia o los problemas que encuentra para realizar la búsqueda (Tenopir, Nahl y Howard 1991; Solomon 1993; Wang, Hawk y Tenopir 2000).

En cuanto a los métodos de análisis, las estadísticas descriptivas están presentes en la mayoría de los trabajos, tanto si se han utilizado datos cuantitativos como si se ha llegado a establecer categorías susceptibles de contabilizar. Los métodos inferenciales se emplean en menos ocasiones siendo los más frecuentes Chi-cuadrado (Hildreth 1997) U Man-Whitney (Sutcliffe 2000) Correlación (Drabenstott 1996a).

Entre los métodos cualitativos el más habitual es el análisis de contenido. Puede aplicarse a las transacciones para crear categorías sobre los términos de búsqueda, los errores o las secuencias del proceso de búsqueda (Drabenstott y Weller 1996d; Bates 1993); o a la información obtenida mediante entrevistas (Spink 1996a), observación o protocolos verbales (Fidel 1999; Slone 2000). En menor medida se han empleado la

técnica de la comparación constante utilizada en la teoría fundamentada (Hert 1997; Connaway, Johnson y Searing 1997; Spink y Saracevic 1997) y la descripción densa o etnográfica (Tenopir, Nahl y Howard 1991, Fidel 1999).

1.4.2 Variables.

En relación con las variables no sólo llama la atención su gran variedad sino la falta de coincidencia, tanto en su definición como en su utilización. Como consecuencia, ha sido necesario interpretar el concepto en un sentido amplio, para dar cabida bajo una misma etiqueta a la multitud de casos que se presentaron.

Se ha considerado como variable el valor (numérico o no) de cualquier observación o registro realizado por una persona o máquina y el deducido o calculado de alguna manera de las observaciones o registros. Estos valores son descriptivos de situaciones, personas, sistemas, acciones realizadas y opiniones emitidas.

El conjunto de las variables utilizadas se ha dividido en cuatro categorías: 1) usuarios, 2) sistema, 3) entorno y proceso de búsqueda y, 4) resultados de la búsqueda.

Las relacionadas con los *usuarios* (Tabla 1.22) representan un conjunto de características individuales que se utilizan como variables independientes con la finalidad de comprobar su influencia en el comportamiento de búsqueda y en la valoración que se hace de los resultados. El objetivo es crear modelos de usuarios basados en las diferencias individuales, modelos que deben permitir diseñar sistemas que se adapten al comportamiento de búsqueda de grupos con atributos diferentes.

Características de los usuarios	
<ul style="list-style-type: none">• Características demográficas<ul style="list-style-type: none">GéneroEdad• Formación<ul style="list-style-type: none">Titulación o formación principalNivel• Nivel profesional• Experiencia<ul style="list-style-type: none">Con el SRICon la base de datosCon los ordenadores en generalCon la búsqueda de información	<ul style="list-style-type: none">• Conocimiento de la materia y del área temática• Modelos mentales• Características personales<ul style="list-style-type: none">InteligenciaEstilo de aprendizajeHabilidad para el razonamiento simbólicoHabilidad lingüísticaCreatividad• Estados emocionales o afectivos

Fig.1.22. Variables. Características de los usuarios

La variable que se ha analizado en más investigaciones es la experiencia con la recuperación de información en general o con el sistema que se está estudiando en concreto (Dalrymple 1992; Hsieh-Yee 1993; Lazonder 2000; Wang, Hawnk y Tenopir 2000). Se ha valorado la influencia de factores demográficos como género o edad (Borgman 1995; Connaway 1995), de características personales como estilo cognitivo o personalidad - medidos por medio de algún test normalizado - (Allen 2000; Borgman

1989; Logan 1990; Ford 2000; Qui 1993), del nivel profesional o académico (Connaway 1997; Ferl y Millsap 1996). Con menor frecuencia se han tenido en cuenta la formación o especialización (Siegfried, Bates y Wilde 1993) y los modelos mentales, entendidos como la idea que el usuario tiene sobre las características, estructura, contenido y funcionamiento del sistema (Borgman 1989; Dimitroff 1992; Silkes 1997). También se ha analizado la influencia de los estados afectivos o emocionales (Kulhthau 1992; Tenopir, Nahl y Howard 1991).

Las variables relacionadas con el *sistema* se utilizan en aquellos estudios que comparan diferentes sistemas de recuperación, el mismo sistema con diferentes bases de datos o la eficacia de diferentes tipos de interfaces para un mismo sistema. Habitualmente se consideran variables independientes y se relacionan con el rendimiento que obtienen los usuarios del sistema y con las características del proceso de búsqueda.

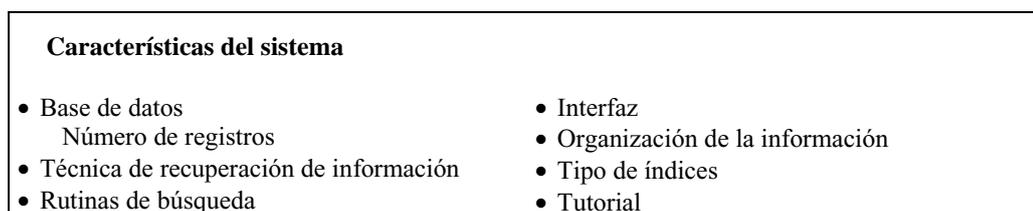


Fig. 1.23. Variables. Características del sistema

Se ha analizado la incidencia de dos elementos clave como son el tamaño de la base de datos (Tillotson 1995, Larson 1991b) y las técnicas de recuperación de información (Efthimiadis 2000; Hancock-Beaulieu 1993; Larson 1992). Otros factores que afectan a la interacción son la forma en que está organizada la información en el sistema (Dimitroff y Wolfram 1995), su presentación en la interfaz (Allen 1998; Beheshi 1996; Large y Bialek; Hildreth 2001; Thomas 2001), los tipos de índices (Spink y Saracevic 1997). También se ha valorado el beneficio que supone la incorporación al sistema de una rutina que seleccione la mejor opción de búsqueda en función del tipo de enunciado utilizado por el usuario (Drabenstott y Weller 1996d) o la influencia de las características de diferentes tipos de ayuda en línea (Blecic 1999).

El contexto en el que surge la necesidad de información y la claridad con que el usuario la percibe y es capaz de expresarla con sus propios términos ha generado un conjunto de variables que hemos agrupado bajo la denominación entorno de búsqueda. Éstas se consideran independientes cuando se quiere comprobar cómo influyen en el desarrollo de este proceso o en la valoración de los resultados, aunque también pueden considerarse dependientes de las características de los usuarios.

Entorno de búsqueda	
<i>Contexto de la búsqueda</i>	<i>Definición del problema</i>
<ul style="list-style-type: none"> • Entorno de trabajo en el que se inserta la búsqueda (área temática, teórico, práctico, etc.) • Finalidad de la búsqueda • Fase de desarrollo del proyecto para el que se busca información • Tiempo que se ha previsto invertir en la búsqueda • Presuposiciones (cantidad de información que se cree que existe sobre el tema de búsqueda) 	<ul style="list-style-type: none"> • Nivel de definición de la necesidad de información • Expresión de la necesidad de información <ul style="list-style-type: none"> Claridad Especificidad Complejidad • Dificultad • Número de búsquedas por necesidad de información expresada

Fig. 1.24 Variables. Entorno de búsqueda

Entre ellas se puede incluir el área de conocimiento para la que se realiza la búsqueda (Ellis, Cox y Hall 1993) o el tipo de trabajo para el que se va a utilizar la información recuperada (Connel 1995; Fidel 1999; Hirsh 1997) o el nivel de definición de la necesidad de información (Hildreth 2001). Éste último puede ser expresado directamente por el usuario o evaluado mediante la diferenciación entre búsquedas por ejemplar conocido (bien definidas) y búsquedas “de documentos desconocidos” (no claramente definidas) (Slone 2000; Wildemuth y O’Neil 1995).

Las agrupadas bajo el epígrafe *proceso de búsqueda* son variables relacionadas con la interacción directa del usuario con el sistema de recuperación y tienen como objetivo reflejar las acciones realizadas durante esta fase y entender qué objetivos las justifican. Habitualmente se utilizan como variables dependientes de las características del usuario

o del sistema e independientes cuando se analiza cómo varía el rendimiento en función de la estrategia de búsqueda.

Algunos factores empleados para caracterizar el proceso de búsqueda y determinar el grado de interacción son: el número de puntos de acceso o de descriptores empleados, el número de pantallas visualizadas, los errores cometidos, los operadores utilizados, los enlaces hipertextuales activados y el tiempo empleado (Barab 1997; Blecic et al. 1999; Cooper 2001; Ferl y Millsap 1996; Hildreth 1997; Irgens 2000; Jansen, Spink y Saracevic 2000).

Proceso de búsqueda	
<ul style="list-style-type: none"> • Puntos de acceso utilizados • Tipo de punto de acceso • Descriptores utilizados <ul style="list-style-type: none"> Número total Tipo de descriptores • Fuente de los descriptores <ul style="list-style-type: none"> Estructuras de conocimiento Otros documentos Retroalimentación automática • Combinaciones de términos de búsqueda <ul style="list-style-type: none"> Número de combinaciones Tipo de operador utilizado • Órdenes y opciones del sistema utilizadas • Movimientos <ul style="list-style-type: none"> Número de movimientos Tipo de movimiento 	<ul style="list-style-type: none"> • Ciclos <ul style="list-style-type: none"> Número de ciclos Tipo de ciclo • Visualización de registros o documentos <ul style="list-style-type: none"> Número de registros visualizados en cada ciclo Número total de registros visualizados • Errores <ul style="list-style-type: none"> Número de errores Tipo de error Circunstancias del tipo de error • Tiempo empleado en la sesión <ul style="list-style-type: none"> Tiempo empleado en la resolución de problemas Tiempo empleado en ejecutar la búsqueda • Reacciones y emociones del usuario durante la búsqueda • Modificación del objetivo inicial de la búsqueda

Fig. 1.25. Variables. Proceso de búsqueda

Cualquier secuencia de acciones consecutivas – elección de un campo, combinación de términos, navegación, visualización y selección de registros – constituye un ciclo de búsqueda que puede caracterizarse por el tipo de acción predominante o por el objetivo que se pretende alcanzar con ellas (Belkin et al. 1996; Fidel 1991; Hirsh 1996). En los trabajos de corte más cualitativo se consideran también las sensaciones de frustración, ansiedad, desorientación o éxito, experimentadas por el usuario durante el proceso de búsqueda (Wang, Hawk y Tenopir 2000; Slone 2000).

Una de las variables que se utiliza con más frecuencia es el tipo de punto de acceso empleado que está relacionado con la definición del problema, afecta a las características del proceso de búsqueda e influye en los resultados (Hildreth 1997; Larson 1991b; Ray y Lang 1997; Slone 2000; Wyly 1996).

Se ha dedicado una atención especial a la búsqueda por materia estudiándose las características de los términos empleados (Bates 1993), su coincidencia con los encabezamientos de materia del catálogo (Carlyle 1989 Drabenstott y Vizine-Goetz 1994; Drabestott y Weller 1996d), la incidencia de los errores ortográficos o tipográficos (Drabenstott y Weller 1996b), los problemas para encontrar términos alternativos cuando hay que reconducir la búsqueda (Connel 1995, Fidel 1999) y las secuencias de movimientos realizados en la modificación (Ferl y Millsap 1996). Además, es habitual la comparación entre el éxito de las búsquedas temáticas

realizadas en el campo materias y las que se plantean como palabras clave (Bleicic 1999; Larson 1991b; Hildreth 1997; Martínez Arellano 1999; Tillotson 1995).

En la mayoría de las investigaciones la búsqueda se entiende de forma autónoma y no se tienen en cuenta sesiones anteriores o posteriores. Sin embargo, en situaciones reales pueden ser necesarias varias sesiones de búsqueda para resolver la misma necesidad de información. Si se entiende el proceso de búsqueda en este sentido amplio, interesa conocer cómo se modifican los términos, las estrategias e incluso los objetivos iniciales (Hert 1997; Spink 1996a).

Finalmente, queda el grupo de variables relacionadas con los *resultados de la búsqueda*.

Rendimiento	
<ul style="list-style-type: none">• Número de registros recuperados	
Eficacia	Eficiencia
<ul style="list-style-type: none">• Juicios de relevancia por registro<ul style="list-style-type: none">Relevancia temáticaRelevancia contextual (<i>Utilidad</i>)• Juicios de relevancia por búsqueda<ul style="list-style-type: none">ExhaustividadIntegridad de los resultados	<ul style="list-style-type: none">• Coste de la búsqueda• Valoración subjetiva del coste de la búsqueda en función de los resultados• Valor de los resultados de búsqueda en función del tiempo empleado• Valor de la búsqueda en su conjunto
<i>Precisión</i>	
Exactitud de los resultados	

Fig. 1.26 Variables. Rendimiento

Las medidas de rendimiento que se han utilizado combinan los dos puntos de vista en la evaluación de sistemas de recuperación de información. Por una parte, se mantiene las medidas tradicionales, exhaustividad y precisión, basadas en la relevancia temática (Lancaster et al. 1994; Sutcliffe, Ennis y Watkinson 2000). Sin embargo, la idea de “búsqueda en situación” introduce un principio de incertidumbre en la valoración de la relevancia que pasa a considerarse una medida dependiente del contexto. Algunos autores emplean el término utilidad, por contraste con el de relevancia, para denominar el criterio empleado para valorar, en este caso, la satisfacción o la integridad y la exactitud de los resultados obtenidos (Hildreth 2001; Jokic 1997; Spink, Greisdorf y Bateman 1998).

1.4.3 La investigación en España.

Como han señalado Delgado López-Cózar (2000) o Moya Anegón (2000), la investigación sobre recuperación de información en España es prácticamente inexistente. Esta afirmación se puede extender al estudio de la búsqueda en catálogos en línea. A pesar de esta evidencia, se ha realizado una revisión de la bibliografía sobre la materia, con el objetivo de completar el panorama internacional.

Al igual que en el apartado anterior, se han seleccionado solo trabajos de investigación original, aparecidos a partir de 1990 en alguna de las publicaciones que se encuentran en el *ranking* elaborado Moya Anegón y Jiménez Contreras (1999)⁵⁶, así como en monografías o publicaciones de asociaciones profesionales. La información se ha organizado siguiendo las mismas divisiones que el epígrafe anterior: técnicas, métodos y variables.

La mayor parte de los trabajos recopilados se realizaron en un momento de transición entre el catálogo manual y el automatizado. Esta coyuntura justifica que el objetivo fundamental fuera conocer el grado de satisfacción de los usuarios con la nueva herramienta de búsqueda y apreciar su nivel de aceptación.

La técnica de recogida de información que se utiliza prácticamente en todas las investigaciones es el cuestionario (Andreu, Escañuela y Torres 1996; Bes Gracia y Romance Carcas; Colomer y Serra, 1992; Pérez Díez 1996). La única excepción es el somero análisis transaccional realizado por Menéndez Llana (2001) en las bibliotecas públicas asturianas. El análisis de los datos obtenidos por cualquiera de estos dos medios es meramente descriptivo, presentándose los resultados mediante gráficos o tablas.

⁵⁶ Las publicaciones son, en este orden: Revista Española de Documentación Científica, Boletín de la Anabad, Documentación de las Ciencias de la Información, IWE. El profesional de la información, Jornadas Catalanas de Documentación Automatizada, Boletín de la Asociación Andaluza de

Respecto a las variables, sólo algunos trabajos pretenden conocer el perfil de los usuarios del centro y, en consecuencia, utilizan variables de las que se han incluido en este grupo: nivel de estudios, ocupación (Colomer y Serra 1992) y experiencia en el uso de ordenadores, catálogos o bases de datos (Pérez Vicente 1996).

Los elementos de sistema que se valoran son: facilidad de consulta, necesidad de ayuda (Colomer Serra 1992), presentación de la información, tiempo de respuesta, secuencia de las pantallas o claridad de los mensajes (Pérez Vicente 1996). No se analiza el entorno búsqueda con precisión pero se considera algún factor como el nivel de definición de la necesidad de información, distinguiendo entre búsqueda por ejemplar conocido o no (Bes Gracia y Romance Carcas 1991).

En cuanto al proceso, apenas se examina su desarrollo aunque sí se tienen en cuenta, por ejemplo, los campos de búsqueda utilizados (Andreu, Escañuela y Torres 1996, Bes Gracia y Romance Carcas 1991). El estudio de Menéndez Llana (2001) contabiliza el número de búsquedas realizadas, visualizaciones, errores y búsquedas nulas durante un periodo de año y medio. Sin embargo, no permite hacerse una idea, ni siquiera general, de las características del proceso: no sólo porque los datos son muy genéricos, sino porque no define qué es un error ni una búsqueda nula.

Como peculiaridad frente a los estudios de otros países, se tiene en cuenta factores de corte totalmente pragmático como la valoración que hacen los usuarios del número de terminales que existen en la biblioteca o sobre su ubicación. La utilización de estas variables se justifican dentro de la valoración global que se intenta hacer del proceso de automatización de un centro concreto.

El rendimiento del sistema no se evalúa teniendo en cuenta la relevancia, temática o contextual, ni la exhaustividad de la respuesta. Se hace una valoración subjetiva, en términos de mayor o menor satisfacción, que juzga tanto los resultados de la búsqueda como la amigabilidad del sistema para la interacción. En estas condiciones los resultados son, en general, satisfactorios y los usuarios se decantan por los OPAC's frente a los catálogos de fichas, aunque los resultados de la búsqueda no sean demasiado pertinentes.

2. Objetivos y Metodología

En el capítulo anterior se han señalado entre las principales causas de los fallos en la búsqueda temática la escasa coincidencia del vocabulario empleado por los usuarios con el utilizado en los índices del sistema y el desconocimiento de la estructura de los encabezamientos de materia. La importancia de este problema se ha confirmado en los estudios que han comprobado el grado de similitud terminológica y sintáctica entre los enunciados de búsqueda y los encabezamientos de materias. Sus resultados demuestran que la coincidencia exacta se produce en menos de la mitad de las búsquedas. Si se aplica algún tipo de normalización y se utiliza una combinación de técnicas de recuperación junto con el índice de materias se puede conseguir que hasta un 40% más de los enunciados, cuya coincidencia con el índice es sólo parcial, recupere información potencialmente relevante.

Como apuntaba Carlyle (1989) aunque estos estudios coinciden en el objetivo, varían metodológicamente en tan gran medida que no son comparables satisfactoriamente. Las diferencias afectan a la definición de búsqueda por materias (sólo en el campo materias o también en el campo título), al tipo de catálogo que se analiza (manual o

automatizado), a las técnicas de recogida de datos (cuestionarios y transacciones) y a las categorías de equiparación.

Los primeros trabajos equiparaban las expresiones de los usuarios con los encabezamientos utilizados en catálogos manuales: Knapp (1944) estudió los de las bibliotecas del Chicago Teachers College y Woodrow Wilson Junior College, Tagliacozzo y Kochen (1977) analizaron los de tres bibliotecas de la Universidad de Michigan y el de la Ann Arbor Public Library. El trabajo exploratorio realizado por Bates (1977) significó el inicio de las investigaciones en las que se comparan los enunciados de búsqueda con el vocabulario utilizado en la *Library of Congress Subject Headings*.

En la década de los 80 se produce un cambio tanto en el entorno analizado como en el método empleado para recopilar los enunciados de búsqueda: se analizan los índices de los catálogos en línea y las expresiones de los usuarios se recogen utilizando el registro de las transacciones de estos sistemas. Trabajos como los de Carlyle, Markey Drabenstott, Hunter, Lester, o Taylor, responden a estas características.

Taylor (1984) utilizó las transacciones para estudiar el grado de coincidencia entre el vocabulario de los usuarios y los registros de autoridad del sistema NOTIS en la Northwestern University. Posteriormente, Lester comprobó la eficacia de diferentes

técnicas de recuperación para mejorar el rendimiento de las búsquedas por materias en el mismo catálogo en función de la coincidencia entre los enunciados de los usuarios y los índices del sistema (Lester 1989, Roose y Lester 1988). Hunter (1991) examinó la incidencia de los errores tipográficos y de la terminología en el éxito de la recuperación por materias en el catálogo de la North Carolina State University. Por su parte, Carlyle (1989) hizo una detallada revisión de la metodología aplicada en investigaciones anteriores y una rigurosa definición de las variables que permiten equiparar el vocabulario y la sintaxis de las búsquedas con los índices de materia. A continuación, comparó una pequeña muestra de los enunciados de búsqueda empleados en la University of California Los Angeles con los encabezamientos de la LCSH.

El estudio de Markey (1984) sirvió como punto de partida para las dos investigaciones más ambiciosas que se han realizado sobre el tema. Se trata de dos proyectos relacionados, y coordinados por la autora, cuyos resultados han aparecido en diversas publicaciones a lo largo de la década de los 90. Los títulos de los proyectos fueron “Enhancing a new subject access design to online catalogs” (financiado por OCLC’s Library and Information Science Research Grant Program) y “Testing a new subject access design to online catalogs” (promovido por el Department of Education’s College Library Technology and Cooperation Grants). En ellos se analizaron enunciados empleados en los OPACs de la Syracuse University, la University of California Los Angeles, la University of Kentucky y la University of Michigan con el

objetivo de conocer sus características y grado de coincidencia con la LCSH. Basándose en los resultados se proponen los “árboles de búsqueda”: diagramas de toma de decisiones que permiten elegir la opción de búsqueda más adecuada en función del tipo de expresión y del número de palabras (Drabenstott 1994, 1996a; Drabenstott y Vizine-Goetz 1990, 1994; Drabenstott. y Weller 1994, 1995, 1996a-d; Vizine-Goetz y Drabenstott 1991).

Estas investigaciones se circunscriben al ámbito anglosajón, a las características de la lengua inglesa y, específicamente, de la LCSH. No hemos encontrado estudios similares en los que se analicen enunciados en castellano para compararlos con los índices de algún catálogo en línea o con los encabezamientos de alguna de las listas que se utilizan en el estado español. La comprobación de este vacío y la evidencia de que los problemas del acceso por materias son generalizables a cualquier idioma fueron los motivos que nos animaron a abordar esta investigación. El trabajo realizado en el conjunto de investigaciones que se acaban de referenciar - y que se irán reseñando a lo largo de este capítulo - han servido de apoyo tanto para la definición los objetivos y las hipótesis de trabajo como para establecer la metodología.

2.1 Objetivos e hipótesis de trabajo.

En este estudio se han analizado las expresiones de búsqueda de los usuarios del OPAC de una biblioteca universitaria y se ha comparado su vocabulario y su sintaxis con el índice de materias. No se pretende, como sucede en otras investigaciones, comprobar si el enunciado del usuario expresa la materia por la que tiene interés, o el número de encabezamientos recuperados, o en qué medida los encabezamientos seleccionados recuperan documentos relevantes. De los dos factores que se han estudiado – características de los enunciados y coincidencia con el índice - depende la eficacia de diferentes técnicas de recuperación que pueden mejorar los resultados de las búsquedas temáticas. Específicamente los objetivos que nos se han planteado han sido:

- ✓ Detectar los errores que aparecen en los enunciados de búsqueda y comprobar su incidencia en la recuperación de información.

- ✓ Conocer las características de los enunciados de búsqueda utilizados por los usuarios en el campo materia: número de palabras, estructura sintáctica, tipo de búsqueda (campo 6XX del formato IBERMARC).

- ✓ Analizar el grado de coincidencia entre la terminología y la sintaxis empleada por los usuarios y las utilizadas en el índice de materias del catálogo.

- ✓ Comprobar si existe alguna relación entre las características de las búsquedas y su coincidencia con los términos del índice.

- ✓ Comprobar si existen diferencias significativas entre las búsquedas realizadas desde terminales ubicados en facultades pertenecientes a diferentes áreas de conocimiento. Se ha verificado la similitud tanto de las características de los enunciados de búsqueda como del grado de coincidencia de éstos con los términos del índice.

Las investigaciones que se han realizado previamente sobre este tema nos han proporcionado las siguientes hipótesis de partida relacionadas con los objetivos que acabamos de exponer:

- ✓ Los errores que se han detectado en las búsquedas en otros análisis han sido:
 - Los usuarios de los catálogos en línea emplean expresiones que, por distintas características formales, pueden considerarse búsquedas “no válidas”

- Aproximadamente un 5% de los enunciados incluyen errores ortográficos o tipográficos.

Estudios que mantienen estas afirmaciones: Blečić (1999: p. 525), Drabentstott (1996: p. 90), Drabentstott y Weller (1996c: p. 117-119), Hunter (1991: 400), Jones (1986: p. 4), Lester¹ (1989: p. 194, 197), Markey (1987: p. 38), Peters (1989: p. 269), Walter (1987: p.76), Walker y Jones (1987: p. 129-131), Zink (1991: p. 53).

✓ El análisis de las características de las búsquedas investigaciones anteriores demuestran que:

- Los usuarios de los catálogos en línea utilizan mayoritariamente enunciados de una o dos palabras para expresar las materias sobre las que necesitan información.
- El tipo de búsqueda más frecuente en el campo materias es la temática *propiamente dicha* (el encabezamiento que le correspondería sería una etiqueta 650 del formato IBERMARC).

Estudios que mantienen estas afirmaciones: Drabentstott (1996a: p. 91), Drabentstott y Vizine-Goetz (1994: p. 157), Hancock-Beaulieu (1987: p. 312), Jones, Cunningham y McNab (1998: p. 293-294), Lester (1989: 176, 186), Walker y Jones (1987: p. 122)

¹ Citado por Drabentstott y Vizine-Goetz (1994).

✓ Respecto a la coincidencia de la terminología de los usuarios y la de los encabezamientos de materia y su relación con las características de las búsquedas se ha demostrado que:

- Menos del 40% de los enunciados de búsqueda coinciden exactamente con los índices temáticos del catálogo.
- Alrededor de un 15% de las expresiones están incluidas en encabezamientos de materia más amplios.
- Si se corrigen los errores tipográficos y ortográficos, se eliminan las diferencias singular/plural, se normalizan las formas abreviadas y las fechas, y se compara sólo el vocabulario - sin tener en cuenta la sintaxis de las expresiones – la coincidencia puede incrementarse hasta un 15%.
- Existe relación entre las características del enunciado (tipo de búsqueda, número de palabras...) y su grado de coincidencia con los índices del sistema.

Estudios que mantienen estas afirmaciones: Carlyle (1989: p. 44, p. 51), Drabentstott (1996a: p. 91), Drabentstott y Vizine Goetz (1994: p. 164, 175, 183, 194), Hunter (1991: p. 399), Larson (1986: p. 127²); Lester (1989: p. 172), Taylor (1984: p.12), Wilkes y Nelson (1995: p. 63)

² Citado por Yee y Shatford (1998).

✓ Sobre las diferencias de las búsquedas por áreas de conocimiento, trabajos anteriores han demostrado que:

- Existen diferencias en la terminología y en el comportamiento de búsqueda entre los usuarios especializados o con formación en áreas de conocimiento distintas.

Estudios que mantienen esta afirmación: Bates, Wilde y Siegfried (1993: p 14-19), Borgman (1986: p. 392; 1989: p. 248), Cool (1993: p. 275), Ellis, Cox y Hall (1993: p. 365-367), Saracevic y Kantor (1988: 195-196, 198-201), Shaw (1995: p.330, 340), Stone (1982: p. 310), Swinger (1992: p. 62), Wiberley (1983: p. 421), Wiberley y Jones (1989: p. 640).

2.2 Metodología.

Las decisiones sobre la metodología empleada para la realización de la investigación están relacionadas con dos aspectos fundamentales que también se utilizan para estructurar este apartado: a) la recopilación y selección de la información y b) las variables y los métodos que se han empleado en el análisis.

2.2.1 Materiales.

2.2.1.1 Recogida y depuración de datos.

Los estudios previos, en los que se han analizado los enunciados de búsqueda por materias, han empleado como técnicas de recogida de datos los cuestionarios, las entrevistas y las transacciones. En este trabajo se ha optado por realizar un análisis transaccional. No sólo es la técnica que más se ha utilizado de las tres desde la aparición de los catálogos en línea, también es la que proporciona información más directa y exhaustiva sobre los errores y las características de las expresiones de los usuarios (Borgman 1996, Frías y Martín Rodríguez 1999, Jones 1997, Peters 1993).

Respecto al centro en el que se iban a recoger los datos, desde el primer momento se optó por una biblioteca universitaria. Sin embargo, la elección de un catálogo concreto estuvo condicionada por dos factores de carácter totalmente pragmático. En primer lugar, el centro tenía que trabajar con un sistema de gestión que incluyera un módulo de grabación de las transacciones que registrara la información que se necesitaba. También era imprescindible que el Servicio de Bibliotecas estuviera dispuesto a facilitárnosla. Después de algunos tropiezos iniciales³, se contactó con el Servicio de

³ Los inconvenientes alegados por el Servicio de Bibliotecas de la Universidad de Salamanca nos obligaron a ponernos en contacto con otros centros. El primero fue la Biblioteca de la Universidad de Vigo donde nos explicaron el funcionamiento del sistema, nos aclararon unas cuantas dudas y nos

Bibliotecas de la Universidad Complutense de Madrid que se mostró dispuesto a colaborar, facilitando información sobre su catálogo (CISNE)⁴.

Desde el año 2000 CISNE se gestiona con el sistema INNOPAC/MILLENIUM, uno de los programas con más implantación internacional en grandes instituciones y consorcios. Facilita acceso a los fondos localizados en 30 bibliotecas de centro con más de 100.000 usuarios. La base de datos contiene información sobre 2.500.000 de monografías, más de 40.000 títulos de publicaciones periódicas y más de 80.000 títulos de materiales no librarios. MILLENIUM cuenta, además de los módulos habituales, con un “generador de informes”⁵ que graba las operaciones realizadas por el sistema y realiza informes estadísticos. Suministra información sobre el terminal desde el que se realizan las consultas; la fecha y hora de las mismas (con intervalos de un minuto), los índices en los que se han efectuado, los términos de búsqueda y el número de registros recuperados.

La información que se nos proporcionó fue un listado de los enunciados de búsqueda utilizados en el campo materia desde todos los terminales entre los meses de marzo a

facilitaron las primeras transacciones. Desgraciadamente no se han podido utilizar en este trabajo porque no identificaban los terminales.

⁴ El nombre del catálogo de la Biblioteca es una forma abreviada del nombre del fundador de la Universidad Complutense, el Cardenal Cisneros. Se puede consultar en línea en el portal web de la BUC <<http://www.ucm.es/BUCM>> [Consulta: 12 de octubre de 2002]

⁵ *Guía de uso del generador de informes estadísticos de Millenium: versión 1.1.* Madrid: Universidad Complutense, Biblioteca, 2001

junio de 2001. En total dispusimos de 34.905 enunciados registrados en tres grabaciones:

Del 6 al 15 de marzo	17.605 búsquedas
Del 30 de marzo al 4 de abril	8.531 búsquedas
Del 28 de mayo al 1 de junio	8.769 búsquedas

El fichero de texto recibido se formateó mediante un programa⁶ diseñado específicamente para tabular los datos con el objetivo de poder manejarlos en una hoja de cálculo (Excel). El resultado fue un listado alfabético en el que cada expresión se identificaba con el código del terminal, la fecha y la hora de búsqueda.

Tabla 2.1 Transacciones campo materias INNOPAC/Millennium

Expresión de búsqueda	Terminal	Hora	Fecha
accion social	500	7:38 PM	Mar 14
accion tutorial	340	5:32 PM	Mar 12
accion y voluntariado	500	2:17 PM	Mar 9
acciones	440	3:55 PM	Mar 7
minimalismo	520	4:05 PM	Mar 7
minimalismo	520	4:06 PM	Mar 7
minimalismo	900	1:37 PM	Mar 6
mitologia griega	520	5:31 PM	Mar 7
mitologia griega	900	1:32 PM	Mar 10
mitologia griega	900	1:32 PM	Mar 10
palacios de escocia	280	12:38 PM	Mar 6
palacios escoceses	280	12:37 PM	Mar 6
palacios escocia s xi	280	2:33 PM	Mar 6
palazuelo	280	11:07 AM	Mar 8
paleocristiano	280	12:08 PM	Mar 8
paleocristiano y bizantino	280	12:57 PM	Mar 15
paleoecologia	900	11:27 AM	Mar 8

⁶ Para elaborar el programa contamos con la colaboración de Emilio Rodríguez Vázquez de Aldana.

En esta relación (Tabla 2.1), los accesos vía telnet realizados a través de los OPAC de los centros se identifican mediante el código de la sucursal a la que pertenecen, mientras que los que se efectúan a través del OPAC-web se identifican con el terminal 900. El listado se empleó para agrupar las búsquedas por áreas de conocimiento y para eliminar los enunciados duplicados.

Distribución por áreas de conocimiento.

La distribución de las búsquedas por terminales para agruparlas después por áreas de conocimiento serviría para comprobar si existen diferencias entre los enunciados empleados por usuarios de centros pertenecientes a diferentes especialidades científicas. Es cierto que el hecho de que una búsqueda se haga desde un terminal no implica necesariamente que haya sido realizada por usuarios de ese centro y por tanto, de esa especialidad. Sin embargo, este es un criterio que se ha aplicado con frecuencia para diferenciar las búsquedas por áreas temáticas. Además, una rápida exploración de nuestro listado de búsquedas por terminales nos confirmó que sólo una proporción insignificante de las expresiones se refería a temas no relacionados con el área de conocimiento en la que se incluía el terminal desde el que se habían realizado.

La distribución por áreas temáticas se ha ajustado a la nomenclatura de la UNESCO para los campos de la Ciencia y la Tecnología⁷. El resto de las búsquedas (26,95%) son las que se realizaron desde el Opac-Web.

Tabla 2.2 Distribución de las búsquedas por áreas de conocimiento

Area de conocimiento	Terminal	Centro	Búsquedas	
HUMANIDADES				
	200	Filosofía	636	1,82 %
	260	Bellas Artes	651	1,87 %
	280	Geografía e Historia	4043	11,58 %
	300/301/303	Filología	2609	7,47 %
CIENCIAS SOCIALES				
	130	Biblioteconomía y Documentación	18	0,05 %
	220	Educación	2431	6,96 %
	340	Psicología	2168	6,21 %
	400	Derecho	1159	3,32 %
	430	Criminología	8	0,02 %
	440/450	Económicas y Empresariales	1278	3,66 %
	480	Ciencias políticas y Sociología	2490	7,13 %
	500	Trabajo social	443	1,27 %
	520	Ciencias de la información	3961	11,35 %
CIENCIAS MEDICAS				
	600	Medicina	612	1,75 %
	630	Farmacia	219	0,63 %
	650	Veterinaria	234	0,67 %
	670	Odontología	133	0,38 %
	690	Enfermería	456	1,31 %
	700	Óptica	104	0,30 %
CIENCIAS EXACTAS Y NATURALES				
	470	Estadística	106	0,30 %
	800	Matemáticas	296	0,85 %
	830	Físicas	119	0,34 %
	850	Geológicas	236	0,68 %
	870	Biológicas	609	1,74 %
	890	Químicas	454	1,30 %

⁷ Consultada en la página web del Plan Nacional de I+D del Ministerio de Ciencia y Tecnología. http://www.mcyt.es/sepct/PLAN_I+D/codigos_unesco/areas.htm (Consulta: 24 de junio de 2002)

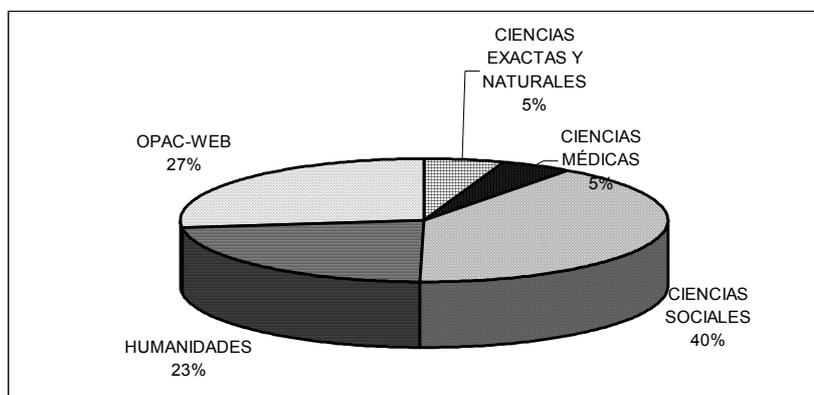


Fig. 2.1 Distribución de las búsquedas por grupos temáticos

Eliminación de duplicados.

Como puede observarse en la tabla 2.1 en el listado original se producían bastantes repeticiones de expresiones de búsqueda que se habían realizado desde el mismo terminal y dentro de un mismo minuto o en intervalos de uno o dos minutos. Esta duplicación era demasiado frecuente para interpretarla como situaciones de “repetición increíble” en las que, según Peters (1989), el usuario repite el planteamiento porque se niega a creer o aceptar los primeros resultados que le presenta el sistema. La escasa diferencia de tiempo nos hizo pensar que se trataba de la misma búsqueda y que la repetición no se producía porque el usuario volviera a teclear el término sino porque utilizaba el navegador para avanzar o retroceder en las pantallas (de hecho casi el 65%

de los duplicados son accesos vía web), o porque seleccionaba el encabezamiento varias veces de alguna presentación.

Antes de seleccionar la muestra de las expresiones que se iban a analizar se filtró la información para eliminar todas aquellas duplicadas y que procedieran del mismo día y terminal con un intervalo de tiempo de dos minutos. El listado se redujo aproximadamente en un 9% (31.782 enunciados)

2.2.1.2 Selección de la muestra.

La muestra seleccionada de este marco debía permitir tanto comparar las características de las búsquedas de cada una de las áreas científicas como obtener resultados generales. El muestreo se ha realizado teniendo en cuenta estas dos condiciones (Cea D'Ancona 1996):

- a) Para obtener resultados comparables se tomó de cada área de conocimiento una muestra aleatoria de 385 enunciados aceptando un riesgo de α de 0,05% con un nivel de confianza del 95,5% para una población infinita y una varianza poblacional estimada del 50% .

b) Para generalizar los resultados se unieron los enunciados seleccionados de los cinco grupos. Como la proporción que correspondía a cada uno de ellos sobre el total de las búsquedas recopiladas era muy diferente, se intentó evitar la distorsión que podía suponer en los resultados de conjunto favorecer a los estratos de menor volumen. Con esta finalidad se ponderaron los enunciados (items) de cada grupo en función del porcentaje que suponían sobre el total⁸. Gracias a esta ponderación se equilibró el valor de los estratos y los resultados son iguales a los de un muestreo aleatorio estratificado por asignación proporcional (aceptando un riesgo de α de 0,05% con un nivel de confianza del 95,5% y una varianza poblacional estimada del 50%).

El tamaño de la muestra se calculó con el programa GRANMO 4.0 del Institut Municipal de Investigació Mèdica de Barcelona y la selección aleatoria se realizó con SPSS 11.0.

2.2.2 Variables.

Las variables se han establecido utilizando tres categorizaciones diferentes: la primera sirve para comprobar la existencia o no de errores, la segunda para conocer las características de los enunciados y la tercera permite examinar el grado de coincidencia de las expresiones de los usuarios con los términos del índice.

2.2.2.1 Errores formales.

Después de consignar para cada búsqueda el área temática a la que pertenecía, se comprobó si existían enunciados de búsqueda no válidos y los errores ortográficos y tipográficos que aparecían.

El Expresiones de búsqueda no válidas

Aunque muchas investigaciones contemplan esta variable, no coinciden los casos que incluyen en ella. Los más habituales son: líneas en blanco, palabras que no se pueden identificar, palabras de contenido erótico, palabras y frases malsonantes, expresiones cortadas o que incluyen algún signo no convencional y secuencias de caracteres sin sentido (Peters 1993). En el caso de las búsquedas por materias en algunos trabajos se

⁸ Recordamos que estos porcentajes son: Ciencias sociales 40%, Humanidades 23%, Ciencias Exacta y Naturales 5%, Ciencias Médicas 5% y OPAC-Web 27%.

han incluido también las que parecen títulos o secuencias autor/título considerándolas búsquedas por ejemplar conocido (Drabenstott y Vizine-Goetz 1994, Drabenstott 1996)

En un primer momento se emplearon las categorías que se han señalado como más habituales, aunque después de haber analizado los enunciados se prescindió de las que no tenían ninguna ocurrencia. También se añadió la categoría “palabras extranjeras” considerándolas “no válidas” porque no pueden recuperar encabezamientos en un índice en castellano y, además, no eran útiles para el análisis que se iba a realizar. No se han eliminado los enunciados que parecían títulos de obras literarias o de películas ni las secuencias autor/título porque pueden ser búsquedas temáticas y formar parte del índice de materias. Finalmente se trabajó con cuatro categorías:

- ✓ **Expresiones en lengua extranjera:** enunciados de búsqueda en una lengua distinta del castellano: *acid rain, food, old indian literature*.

- ✓ **Secuencias aleatorias de letras:** *cbdsijrh, rrrr, qjjkjjfke*

- ✓ **Enunciados de búsqueda cortados** (palabras o frases incompletas) que no permiten deducir lo que el usuario busca: *culturas in , comunicación en , enfermedades neu.*

- ✓ **Enunciados que contienen signos no convencionales:** *vampirj, direct--and directorios.*

En algunos estudios estas búsquedas se eliminan y no se tienen en cuenta en los porcentajes finales. Sin embargo, en éste se ha considerado que son un factor más que afecta al proceso de búsqueda y, como tal, no debería ser ignorado cuando este se analiza. Por tanto los porcentajes, tanto generales como por sectores, se han calculado sobre el total, incluyendo estas expresiones.

E.2 Errores ortográficos y tipográficos

Se trata de detectar los términos incorrectos incluidos por los usuarios en los enunciados de búsqueda. Son errores que se producen por la falta de exactitud en el ingreso de los datos, tanto palabras completas como caracteres. Pueden subdividirse en ortográficos y tipográficos (o mecanográficos), aunque en muchas ocasiones son difíciles de diferenciar y en la bibliografía a veces no se distinguen.

Para comprobar la incidencia de estos errores sobre la recuperación de información en nuestra muestra, estas búsquedas se realizaron dos veces: respetando la grafía de los enunciados y corregidas.

E2.1. Errores ortográficos

La expresión del usuario se diferencia del encabezamiento porque presenta una ortografía incorrecta. Para localizarlos se compararon los datos con la información de la base de datos y con un diccionario de términos correctos. No se han tenido en cuenta algunos signos, como los de acentuación, los signos diacríticos y el uso de mayúsculas, que no plantean problemas en la recuperación porque no son tenidos en cuenta por la mayoría de los sistemas⁹.

B: “inflacción”

I: “Inflación”

E2.2 Errores tipográficos

Se han establecido diferentes categorías para clasificar estos errores, especialmente desde la perspectiva de la evaluación de calidad de las bases de datos. Sin embargo, como indica Randall (1999), los errores tipográficos no son solamente producto de la era de los ordenadores, de hecho existen desde que los manuscritos se copiaban en los monasterios. Alfred Watts, un corrector de pruebas, escribió una obra clásica sobre los errores tipográficos en 1883 en la que los clasificaba en tres categorías: errores de omisión, sustitución y duplicación.

⁹ En el caso de INNOPAC Millenium, las diferencias mayúsculas/minúsculas ni siquiera se registran en las transacciones.

Ortego y Bonal (1996) señalan que en el análisis de bases de datos tradicionalmente se han considerado los cuatro tipos establecidos por Damerau (1964) a mediados de los 60: omisión, inserción, sustitución y permutación. Gardner (1992) especificó las categorías de sustitución y omisión “añadiendo” dos categorías adicionales: errores por la transposición de letras y errores por la inserción de letras.

Nosotros hemos utilizado las cuatro categorías tradicionales que además son las empleadas en estudios similares como el de Carlyle (1989) y Drabenstott y Vizine-Goetz (1994):

E2.2.1 Permutación. Un error de permutación ocurre cuando se intercambian dos letras adyacentes

Aguas residuales

E2.2.2 Sustitución. La búsqueda del usuario se diferencia del encabezamiento porque una letra se ha sustituido incorrectamente por otra.

Economía aplicada

E2.2.3 Inserción. La expresión del usuario se diferencia del encabezamiento porque se ha añadido incorrectamente algún carácter en los términos de búsqueda.

Educaución

E2.2.4 Omisión. La expresión del usuario se diferencia del encabezamiento porque se ha omitido incorrectamente algún carácter en los términos de búsqueda o algún espacio en blanco entre palabras.

Química orgánica

2.2.2.2 Características de los enunciados.

Después de haber comprobado la existencia o no de errores se analizaron las características formales de los enunciados: número de palabras, tipo de búsqueda en el campo materias y estructura sintáctica.

C1 Número de palabras por enunciado.

Siguiendo el modelo de los trabajos de Drabenstott y Vizine-Goetz (1994) o Drabenstott (1996a) para catálogos en línea y el de Jansen, Spink y Saracevic (2000) para servicios de búsqueda en Internet, se han contabilizado todas las palabras de los enunciados de búsqueda, incluyendo las consideradas tradicionalmente palabras vacías. Por tanto por “palabra” debe entenderse cualquier secuencia de caracteres alfabéticos,

numéricos o alfanuméricos precedida y seguida de un espacio en blanco (excepto la primera palabra del enunciado, evidentemente)

C2 Tipo de búsqueda en el campo materia

La tipología de las búsquedas en el campo materias utilizada en los estudios de Drabentstott y Vizine-Goetz (1994), Drabentstott (1996a) o Bates (1993) es muy poco homogénea. Sin embargo, todas ellas combinan las opciones establecidas en los campos 6XX del formato IBERMARC que son las que finalmente se han empleado:

600 Nombre propio de persona ¹⁰	650 Materias
610 Nombre propio de entidad	651 Nombres propios geográficos
630 Títulos uniformes de obras literarias o películas	

C3 Estructura sintáctica.

La categorización de la estructura sintáctica resultó más problemática debido a la multitud de agrupaciones diferentes que se encontraron en los enunciados. Finalmente, para reducir el número de casos y establecer las categorías de análisis, se tomó una muestra aleatoria de 385 búsquedas y se analizaron todas las estructuras que se presentaban. Sólo se mantuvieron como categorías de análisis aquellas que superaban el 1%, las demás pasaron a formar parte del grupo Otras (Otras $\leq 1\%$). Las categorías resultantes fueron:

Nombre común
Nombre común + adjetivo
Nombre común + adjetivo + adjetivo
Nombre común + adjetivo + nombre común
Nombre común + conjunción + nombre común
Nombre común + nombre común
Nombre común + nombre común + adjetivo
Nombre común + nombre propio
Nombre común + preposición + artículo + nombre común
Nombre común + preposición + nombre común
Nombre común + preposición + nombre común + adjetivo
Nombre común + preposición + nombre propio
Nombre propio
Nombre propio + nombre común
Otras

2.2.2.3 Comparación con el índice.

Este último grupo de variables sirve para comprobar el nivel de coincidencia entre las expresiones de búsqueda y los términos del índice de materias del catálogo CISNE. No se ha realizado la comparación con un repertorio de materias de referencia, semejante a la LCSH, puesto que no existe una obra similar para la lengua española que cubra las necesidades de las bibliotecas universitarias. Ante la evidencia de este vacío, muchos centros han ido desarrollando sus propios repertorios en función de sus necesidades

¹⁰ Incluye nombres de personajes de ficción, nombres de familia, etc.

temáticas y lingüísticas, recurriendo para ello a fuentes muy diversas. Es el caso de la Biblioteca de la Universidad Complutense donde se utilizan encabezamientos de materia propios, establecidos por una Comisión en la que participan las Bibliotecas de centro y la Unidad de Proceso y Normalización¹¹.

En cuanto a las variables empleadas en el análisis, hay que insistir en la falta de uniformidad de las investigaciones previas que, en ocasiones afecta al propio objeto de análisis. En algunos casos se confunde la coincidencia con los términos del índice con la relevancia de los términos empleados por el usuario para representar su necesidad de información (Knapp 1944) o se mezclan las características de las búsquedas con la equiparación en una misma categoría (Markey 1984).

Los estudios que establecen categorías más claras y comparables son el de Carlyle (1989) y las investigaciones de Drabenstott con Vizine-Goetz (1994) y con Weller (1994). Aunque se han introducido algunas modificaciones, las categorías que se han establecido son similares a las de estos trabajos ya que eran las más afines a nuestros objetivos y, además, permitirían realizar comparaciones de los resultados.

¹¹ *Encabezamientos de materia: versión 3.0*. Madrid: Universidad Complutense, Biblioteca, 1997

Las variables que miden la coincidencia entre las expresiones de los usuarios y los encabezamientos de materia utilizados en el índice del catálogo CISNE se han establecido utilizando dos criterios:

- ✓ *El grado de similitud terminológica.* En función de este criterio se han creado tres grupos de variables:

Sa. El encabezamiento *coincide* con los términos de búsqueda utilizados por el usuario

$$\mathbf{B = E}$$

Sb. El encabezamiento *contiene todos* los términos de búsqueda utilizados por el usuario

$$\mathbf{B \subseteq E}$$

Sc. El encabezamiento *incluye alguno* de los términos empleados por el usuario.

$$\mathbf{\exists e \in B \mid e \in E}$$

- ✓ *El grado de similitud sintáctica* Este factor ha servido para establecer subgrupos dentro de las categorías anteriores en función de que la

coincidencia terminológica se produjera sin cambios o con cambios en la sintaxis. Se han considerado las siguientes posibilidades:

- La expresión de búsqueda es exactamente igual al encabezamiento en terminología y sintaxis.

- El orden de las palabras del enunciado de búsqueda y el del encabezamiento es diferente. Este cambio de orden puede deberse a:
 - a) No hay coincidencia con la estructura de los encabezamientos de materia, con la sintaxis que establece la precoordinación de los subcampos.
 - b) Las palabras utilizadas aparecen en distinto orden en el encabezamiento sin que tampoco haya relación con el orden de los subcampos.

- Coinciden las palabras significativas pero no los nexos que aparecen en el enunciado de búsqueda.

Las variables definidas de acuerdo a estos dos criterios representan por tanto diferentes grados de coincidencia en función del vocabulario y la sintaxis. Aunque el

orden de las categorías puede parecer semántico en la medida en que las expresiones de la primera categoría tienen más posibilidades de estar próximas en significado que las de las últimas, la ordenación y la similitud no está semánticamente pensada. El conjunto de variables utilizadas y el orden establecido en la secuencia de búsqueda es el siguiente:

Sa. *Coincidencia exacta.* Algún encabezamiento del índice coincide exactamente con la terminología empleada por el usuario en la expresión de búsqueda.

Sa1 *Coincidencia directa.* La expresión del usuario coincide exactamente con un encabezamiento de materia en terminología y sintaxis.

B: Francia situación económica

I: Francia – Situación económica

Sa2 *Cambios en la sintaxis.* Coincide la terminología pero existen diferencias sintácticas.

Sa2.1 *Cambio en el orden de los términos.*

Sa2.1.1 Cambio que coincide con la estructura del encabezamiento en subcampos.

B: Galicia industria

I: Industria – Galicia

Sa2.1.2 Cambio en el orden de las palabras.

B: Dieteticos alimentos

I: Alimentos dietéticos

Sa2.2 *Coincidencia las palabras significativas.* El encabezamiento coincide sólo con las palabras significativas de la expresión del usuario aunque el orden puede variar.

Sa2.2.1 Orden directo

B: Prensa en España

I: Prensa – España

Sa2.2.2 Cambio de orden

Sa2.2.2.1 Subcampos

B: Historia de la música

I: Música - Historia

Sa2.2.2.2 Palabras

B: Carlos Luis de Ribera

I: Ribera, Carlos Luís

Sb. *Coincidencia parcial.* El encabezamiento contiene las palabras del enunciado de búsqueda pero además contiene otros términos adicionales.

Cuando las categorías coinciden con las empleadas en la variable anterior se incluyen los ejemplos pero no vuelven a explicarse para evitar la redundancia.

Sb1 Coincidencia directa

B: Francia situación económica

I: Francia – Situación económica – 1930-1950

Sb2 Cambios en la sintaxis

Sb2.1 Cambio en el orden de las palabras.

Sb.2.1.1 Subcampos:

B: Diccionarios civilización

I: Civilización romana – Diccionarios – Francés

Sb2.1.2 Palabras:

B: Henry James

I: James, Henry, 1843-1916 – Lenguaje – Tesis

Sb2.2 Coincidencia de las palabras significativas.

Sb2.2.1 Orden directo

B: Estructura social en Francia

I: Estructura social – Francia – Grenoble – s. 19

Sb2.2.2 Cambio de orden

Sb2.2.2.1 Subcampos

B: Trastornos de la alimentación

I: Adolescentes – Alimentación – Trastornos

Sb2.2.2.2 Palabras

B: Dominicanas en Madrid

I: Mujeres dominicanas – Madrid (Comunidad autónoma)

Sc. Coincidencia de alguna palabra. El índice incluye alguno de los términos empleados por el usuario. El enunciado de búsqueda contiene una o varias palabras diferentes a las que aparecen en los encabezamientos del índice de materias o contiene términos que se encuentran en diferentes encabezamientos. En este caso las categorías son diferentes a los grupos anteriores por diferentes motivos.

En primer lugar, la propia definición de la variable demuestra que su objetivo es valorar simplemente la coincidencia del vocabulario, no la coincidencia de la sintaxis. De hecho, para realizar la comparación sólo se han tenido en cuenta los términos con significado léxico, eliminando las palabras vacías¹² de las expresiones. Además, era necesario cuantificar el nivel de coincidencia terminológica entre la búsqueda y el índice que estaba implícito en la descripción de las variables precedentes.

¹² Se han considerado palabras vacías los determinantes y las preposiciones.

Sc1 Número de palabras del enunciado presentes en el índice. Número de palabras de la expresión de búsqueda que coincide con algún término de cualquier encabezamiento de materia de la base de datos. Se han considerado las siguientes posibilidades:

Sc1.1 Palabras en el mismo encabezamiento. Como la posibilidad de que todas las palabras del enunciado se encuentren en el mismo encabezamiento ya se considera en las categorías anteriores, aquí se ha valorado el número de términos de la expresión que se encuentran en el mismo encabezamiento de materia.

B: Nuevas tecnologías aplicadas a la educación

I: Nuevas tecnologías - Aspectos sociales

Sc1.2 Palabras en el mismo registro. Coincidencia del enunciado con varios encabezamientos asignados a un mismo registro:

B: Financiación autonómica

R: Comunidades autónomas – Financiación

Hacienda autonómica - España

Las categorías establecidas no sólo evidencian diferentes niveles de coincidencia léxica y sintáctica, también tienen relación con distintos tipos de índices y técnicas de recuperación (Tabla 2.3). De esta manera, los resultados del análisis anterior pueden servirnos para establecer una aproximación sobre qué técnicas resultarían más adecuadas para mejorar el rendimiento del OPAC.

La mayoría de las técnicas de indización y búsqueda que se han examinado se utilizan en sistemas comerciales aunque no todas están presentes en todos ellos. Por otra parte, las técnicas de equiparación parcial no se emplean habitualmente en los catálogos en línea y sólo están presentes en algunos casos en módulos que no están integrados en el “paquete básico”¹³.

En la equiparación de los enunciados con el índice, siguiendo también el modelo de los trabajos de Drabenstott (1996), Drabenstott y Vizine-Goetz (1994) y Carlyle (1989) se tuvieron en algunos aspectos formales que podían incidir en la falta de coincidencia. Según Carlyle (1989) sirven para demostrar con qué frecuencia los truncamientos, los algoritmos de lematización o un mayor número de referencias cruzadas, podrían mejorar los resultados de búsqueda en un sistema en línea.

¹³ Por ejemplo, INNOPAC/Millennium ofrece el motor de recuperación de Altavista como una opción que se puede integrar al sistema de recuperación por palabras clave.

Tabla 2.3 Tipos de coincidencia/Técnicas de recuperación/Índices

Variable	Coincidencia	Búsqueda	Índices
	Coincidencia exacta		
Sa1	Directa	Lectura izquierda-derecha	Encabezamientos en orden directo
Sa2.1.1	Cambio de orden. Subcampos	Lectura izquierda-derecha	Encabezamientos con permutación de subcampos
Sa2.1.2	Cambio de orden. Palabras clave	Palabras clave	Palabras incluidas en los encabezamientos
Sa2.2.1	Palabras significativas. Directa	Lectura izquierda derecha Eliminación de palabras vacías	Encabezamientos en orden directo
Sa2.2.2.1	Palabras significativas. Subcampos	Lectura izquierda-derecha Eliminación de palabras vacías	Encabezamientos con permutación de subcampos
Sa2.2.2.2	Palabras significativas. Palabras	Palabras clave Eliminación de palabras vacías	Palabras incluidas en los encabezamientos
	Coincidencia parcial		
Sb1	Directa	Lectura izquierda-derecha Truncamiento derecha	Encabezamientos en orden directo
Sb2.1.1	Cambio de orden. Subcampos	Lectura izquierda-derecha Truncamiento derecha	Encabezamientos con permutación de subcampos
Sb2.1.2	Cambio de orden. Palabras clave	Palabras clave	Palabras incluidas en los encabezamientos
Sb2.2.1	Palabras significativas. Directa	Lectura izquierda derecha Eliminación de palabras vacías Truncamiento derecha	Encabezamientos en orden directo
Sb2.2.2.1	Palabras significativas. Subcampos	Lectura izquierda-derecha Eliminación de palabras vacías Truncamiento derecha	Encabezamientos con permutación de subcampos
Sb2.2.2.2	Palabras significativas. Palabras	Palabras clave Eliminación de palabras vacías	Palabras incluidas en los encabezamientos
	Coincidencia palabras		
Sc1.1	Palabras	Técnicas de equiparación parcial Eliminación de palabras vacías	Palabras incluidas en los encabezamientos
Sc1.2	Palabras en encabezamientos múltiples	Palabras clave Eliminación de palabras vacías	Palabras incluidas en los encabezamientos

Los enunciados que presentaban este tipo de diferencias se buscaron dos veces – sin corregir y corregidos - para comprobar su influencia en la recuperación de información.

VI. Variaciones singular/plural.

El encabezamiento asignado se diferencia del enunciado de búsqueda en que uno contiene una palabra o palabras en plural (“-s”; “-es”) y el otro contiene la misma palabra o palabras en singular. Esta categoría se puede subdividir en dos:

VI.1 Búsqueda en singular/ Encabezamiento en plural. El encabezamiento contiene en plural (“-s” o “-es”) la misma palabra que aparece en la expresión en singular.

B: Biografía

I: Biografías

VI.2 Búsqueda en plural / Encabezamiento en singular. El encabezamiento contiene en singular la misma palabra que aparece en la expresión en plural (“-s” o “-es”)

B: Mapa geológico

I: Mapas geológicos

V.2 Variaciones en abreviaturas. La búsqueda del usuario contiene una abreviatura de una palabra del encabezamiento o viceversa

B: ue

I: Unión Europea

V3. Variaciones en las fechas o numéricas. El encabezamiento contiene una fecha que se diferencia formalmente de la fecha planteada por el usuario.

B: siglo 19

I: S. 19

2.2.3 Análisis.

En primer lugar se analizaron manualmente las características de los enunciados de búsqueda. Posteriormente, para comprobar la coincidencia de las expresiones con los términos del índice, se aprovecharon diferentes opciones del sistema INNOPAC/Millennium:

- Búsqueda en el campo materias: La búsqueda en esta opción se realiza sobre un índice de encabezamientos de materias con permutación de subcampos y truncamiento a la derecha.

- Búsqueda por palabras clave: Dentro de esta opción se seleccionó la búsqueda en el campo materias.

En los grupos **Sa** y **Sb**, las expresiones se incluyeron en la primera categoría con la que coincidían y sólo cuando no se producía esta coincidencia se pasó a la categoría siguiente. Sin embargo, en el grupo **Sc** se analizaron todas las posibilidades.

Respecto al estudio estadístico, en primer lugar se realizó un análisis detallado de cada variable calculando su distribución de frecuencias. Este análisis nos ha permitido conocer la distribución de las características de los enunciados, la incidencia de los diferentes tipos de errores y el grado de coincidencia entre las expresiones de los usuarios y el índice de materias.

También se han cruzado las variables que permitían confirmar las hipótesis de trabajo que ya hemos señalado. En síntesis:

- Si existe relación entre las características de los enunciados y su nivel de coincidencia con el índice.
- Si existen diferencias significativas entre las características de los enunciados de búsqueda de las diferentes áreas temáticas.

Para verificar estas relaciones se ha utilizado el análisis de la varianza (ANOVA) cuando las variables eran numéricas y el test de Chi-cuadrado para variables categóricas. En ambos casos el umbral de significación se estableció en $< .05$.

3. Resultados

A continuación se exponen los resultados obtenidos del análisis de las variables. Se han organizado siguiendo la división empleada para presentar éstas en la metodología: errores formales, características de los enunciados y similitud con el índice. En cada apartado se presentan en primer lugar los resultados generales y posteriormente los de cada uno de los grupos temáticos que se han establecido. Recordamos que para obtener los resultados generales se han ponderado los casos de cada grupo utilizando el valor del porcentaje que representaba cada uno sobre el total de las búsquedas, esto puede explicar las diferencias que se observarán entre las cifras totales y las de los grupos.

3.1 Errores formales.

Del total de los enunciados analizados el 5,2 % coinciden con alguna de las categorías de la variable *enunciados de búsqueda no válidos*. Más de la mitad son expresiones en alguna lengua extranjera entre las que predomina claramente la lengua inglesa (76%), aunque también hay algunas en francés y alemán. La otra mitad de las expresiones se distribuye entre secuencias aleatorias de caracteres

(25,25%), palabras o frases incompletas (19,19%) y búsquedas que contienen signos no convencionales (5,05%).

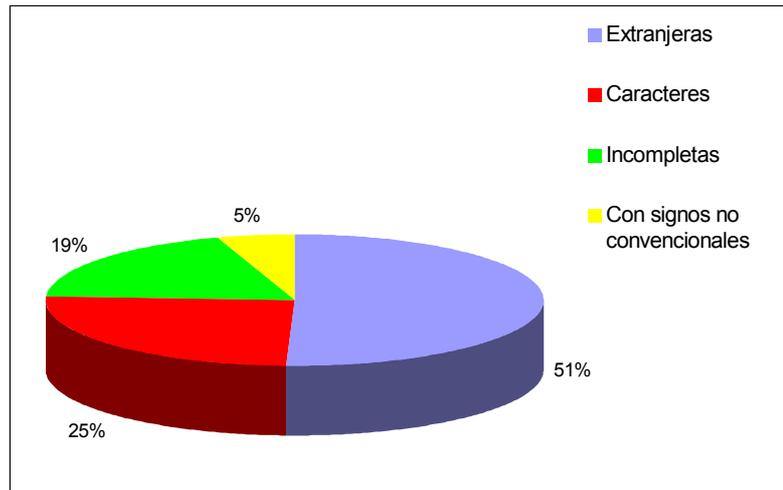


Fig 3.1 Enunciados de búsqueda no válidos: tipos

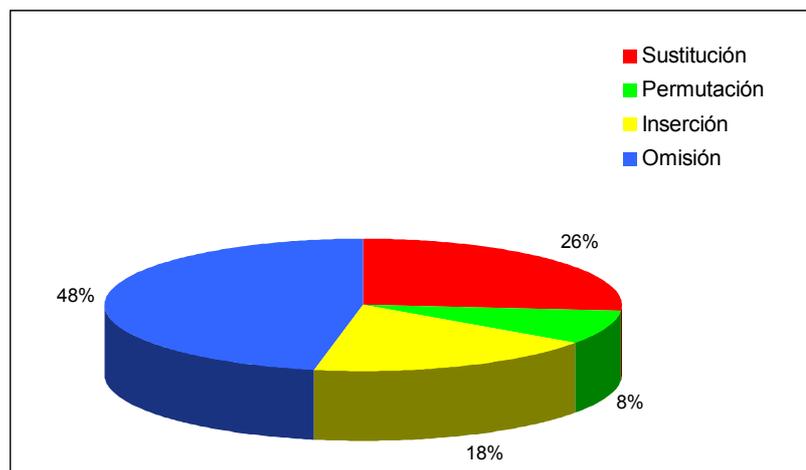


Fig. 3.2. Errores tipográficos: tipos

La proporción de *errores ortográficos* es insignificante (0,5%) mientras que los *tipográficos* afectan al 5,4% de los enunciados de búsqueda. La distribución por tipos de estos últimos (Fig. 3.2) demuestra el predominio de los errores de omisión de los cuales más de la mitad (51%) son omisiones de espacios en blanco. Además, el 16% de los enunciados que incluían errores tipográficos tenían más de uno.

Una primera aproximación a la distribución de estos errores por tipos de búsqueda nos indica que la mayoría se encuentran en búsquedas temáticas (77,5%) de una (26%) o dos palabras (37,3%), en las que se utilizan nombres comunes (21%) o expresiones de nombre común y adjetivo (22,9%) (Tabla A.1 Anexo). Sin embargo estas proporciones se deben a que estas características representan, con gran diferencia, el mayor porcentaje de los enunciados. Si comprobamos qué porcentaje de las búsquedas de cada característica contiene errores tipográficos obtenemos cifras que pueden resultar más reveladoras.

Palabras	Errores	Tipo	Errores	Estructura	Errores
7	34,1	Título	23,46	Nombre común+Adjetivo+Nombre común	15,65
4	9,65	Entidad	10,75	Nombre común+Preposición+Nombre propio	9,91
6	8,79	Persona	9,05	Nombre común+Preposición+Nombre común+Adjetivo	9,9
5	7,91	Lugar	8,8	Nombre propio	9,38
2	6,74	Materia	5,02	Nombre propio+Nombre común	9,28
3	5,96			Nombre común+Preposición+Nombre común	8,15
1	3,61			Nombre común+Nombre común	7,48
8	0			Nombre propio+Nombre común+Adjetivo	7,02
9	0			Nombre común+Adjetivo	6,07
				Nombre común	3,41
				Otras estructuras	< 3

Analizándolos de esta manera, los enunciados que tienen más posibilidades de contener errores tipográficos son los que incluyen mayor número de palabras y los que tienen una estructura sintáctica más compleja, circunstancia que está relacionada, por ejemplo, con la frecuencia de errores de omisión. Por otro lado, también son frecuentes en las búsquedas que contienen nombres propios de entidades (*tricunal económico administrativo, useo de arte reina sofía, archivo genearl de simancas*) o personas (*pompidu, winton churchill*). En este último caso algunas veces difícil saber si el error es tipográfico o se debe a un desconocimiento de la forma correcta del nombre.

En cuanto a la *distribución de los errores entre las áreas*, en primer lugar, como puede apreciarse en la Fig. 3.3, el número de *enunciados no válidos* se reparte de forma bastante homogénea entre los grupos. A pesar de que es ligeramente superior en las búsquedas desde el OPAC-Web, las diferencias que pueden observarse no son significativas (Tabla A.3 Anexo)

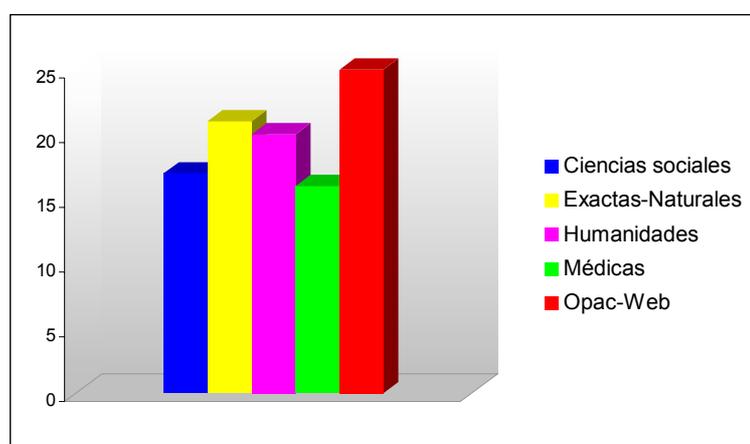


Fig. 3.3 Enunciados no válidos: distribución por áreas

La distribución por tipos (Tabla 3.2) en cada área presenta similitudes respecto a lo que hemos visto en los resultados generales pero también algunas diferencias.

Tabla 3.2 Búsquedas no válidas: tipos por áreas

	Extranjeras	Caracteres	Incompletas	Signos
Ciencias sociales	47,06%	41,18%	11,76%	
Exactas-Naturales	61%	23,81%	9,52%	4,76%
Humanidades	45%	25%	20%	10%
Médicas	25%	43,75%	31,25%	
Opac-Web	64%	4%	24%	8%

Las búsquedas en lengua extranjera son las más numerosas en todos los grupos, excepto en la Ciencias Médicas donde predominan las secuencias aleatorias de caracteres. La proporción más elevada de expresiones en otros idiomas corresponde al OPAC-Web, mientras que en este mismo grupo las secuencias aleatorias de caracteres representan una proporción mínima frente a la que encontramos en el resto de las áreas¹.

El escaso número de *errores ortográficos* también se distribuye de manera uniforme entre los grupos (2 o 3), excepto en el OPAC-Web donde no existen. Respecto a los *errores tipográficos* quizá el rasgo más destacable sea lo poco que se producen en las búsquedas de Ciencias Médicas (Tabla A.5 Anexo). De hecho, las diferencias que se han encontrado sólo son significativas entre este grupo y los de Ciencias sociales ($p=.003$) y Ciencias Exactas-Naturales ($p=.017$).

¹ De hecho, sólo hay diferencias significativas entre los grupos en esta categoría ($p=.025$)

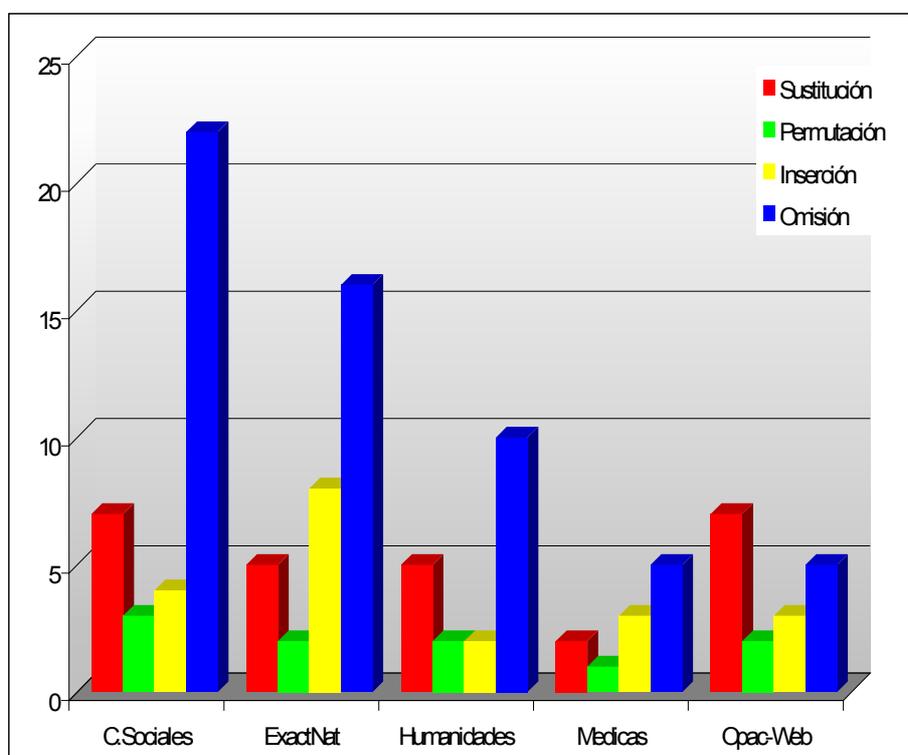


Fig. 3.4 Errores tipográficos: distribución de tipos por áreas

El tipo de error menos frecuente en todas las áreas es el de permutación y el más reiterado el de omisión, excepto en el OPAC-Web donde predominan los de sustitución. Como puede apreciarse en el gráfico, la distribución de las otras categorías no es muy homogénea.

3.2 Características de los enunciados.

El rango del *número de palabras* por enunciado de búsqueda varía entre una y nueve. Sin embargo, la media es $2,10^2$ lo que pone de manifiesto un claro predominio de las búsquedas con una o dos palabras que en conjunto suman un 67,43 % del total.

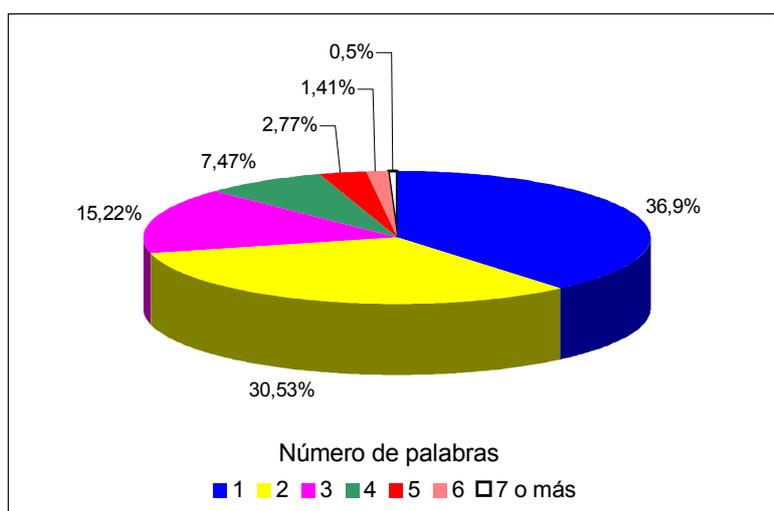


Fig. 3.5 Número de palabras por búsqueda

En cuanto a la distribución por *tipos de búsqueda*, destaca la preeminencia de las búsquedas temáticas, frente a las que tienen como elemento inicial un nombre de persona, lugar o entidad, o bien se refieren a títulos de obras literarias o películas.

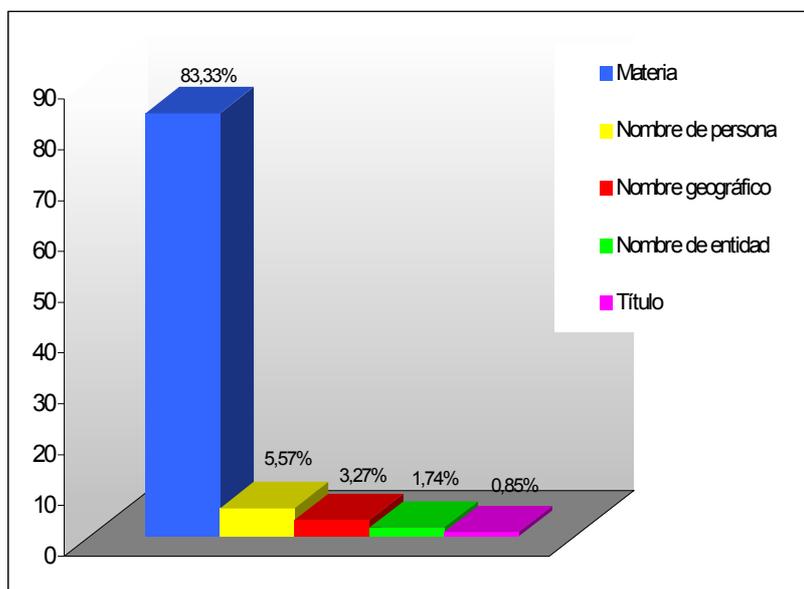


Fig. 3.6 Tipos de búsquedas

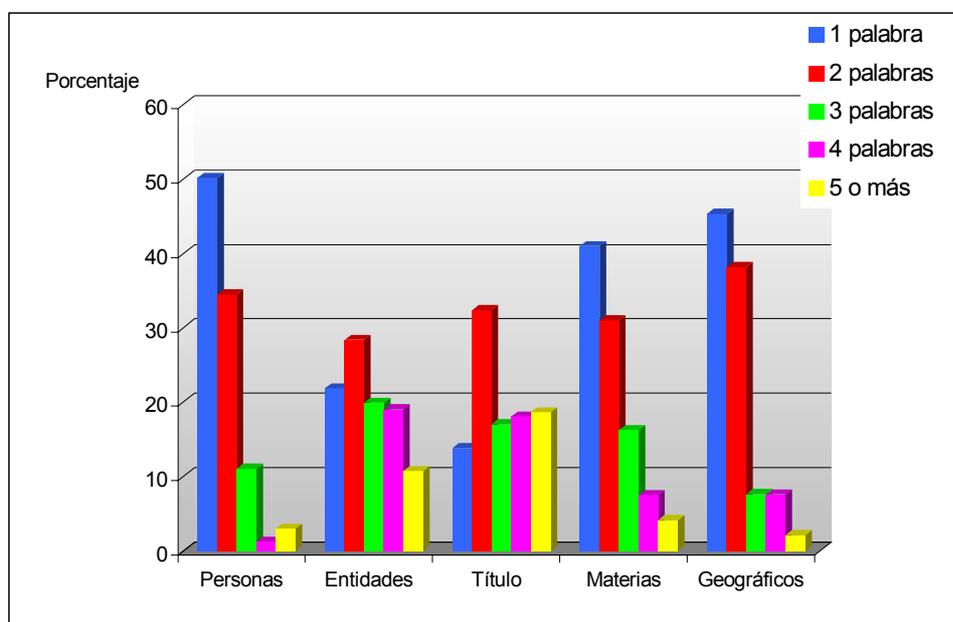


Fig. 3.7 Número de palabras/tipo de búsquedas

² Desv. tip 1,23.

Como puede observarse en el gráfico de la Fig. 3.7, en las búsquedas sobre materias, autor y nombres geográficos los usuarios emplean preferentemente una o dos palabras. No obstante, cuando se trata de títulos o nombres de entidades se utilizan mayoritariamente dos o más. Los resultados están relacionados con el número de términos que habitualmente se utilizan para referirse a estos conceptos: denominar una entidad necesita habitualmente más de un término mientras que es frecuente que sólo se necesiten uno o dos para referirse a un lugar. Estas cifras también indican que es poco habitual que los usuarios incluyan en sus búsquedas especificaciones que concreten el significado de estos términos generales.

<i>Biblioteca Nacional</i>	<i>Tajo</i>
<i>Banco Central Europeo</i>	<i>Chile</i>
<i>La importancia de llamarse Ernesto</i>	<i>Reino Unido</i>

Respecto a los nombres de personas, se emplea mayoritariamente sólo un apellido (50,17%) y cuando se utilizan nombre y apellido, el 50% de las veces el elemento inicial es el nombre de pila. El resto de los casos de más de una palabra en esta categoría corresponde a los que contienen más de un apellido y a los que añaden al nombre de persona una especificación de materia.

<i>Borges</i>	<i>Charlotte Bronte</i>	<i>López Pinciano Alonso crítica e interpretación</i>
<i>Baroja</i>	<i>Víctor López López</i>	<i>Largo Caballero</i>
<i>Proust</i>	<i>Henry James</i>	<i>Gabriela Mistral comentario a su poesía</i>

La mayoría de las búsquedas de materias también tienen una (40,99%) o dos palabras (31,04%). Cuando se utilizan tres (16,26%), el 61,5% de las ocasiones sólo dos son significativas porque una de ellas es un nexos. Este carácter poco específico queda confirmado porque además casi una cuarta parte de las expresiones de uno y de dos términos (23,8% y 23,9% respectivamente) son nombres de disciplinas y una proporción similar de las de tres (23,2%) empieza por un nombre de disciplina.

<i>Historia de Inglaterra</i>	<i>Psicología social</i>
<i>Botánica</i>	<i>Paleontología</i>
<i>Bioquímica</i>	<i>Medicina legal</i>
<i>Informática</i>	<i>Geología de Madrid</i>
<i>Antropología social</i>	<i>Anatomía</i>

En el caso de la estructura, pese a la variedad que se constató en un principio, que obligó a reducir las categorías de análisis, los resultados finales demuestran el predominio de cuatro clases que representan más de dos tercios de los casos. Se trata de las que corresponden a las construcciones sintácticas más simples, lo que resulta lógico si tenemos en cuenta el número de palabras que se contienen los enunciados.

Tabla 3.3 Tipo de estructura de los enunciados de búsqueda

Estructura	Porcentaje
<i>Nombre común</i>	28,92
<i>Nombre común + adjetivo</i>	20,67
<i>Nombre propio</i>	13,05
<i>Nombre común + preposición + nombre común</i>	4,51
<i>*Nombre común + nombre común</i>	3,12
<i>Nombre común + preposición + artículo + nombre común</i>	2,70
<i>Nombre común + preposición + nombre común + adjetivo</i>	1,88
<i>*Nombre común + nombre propio</i>	1,74
<i>Nombre común + preposición + nombre propio</i>	1,74
<i>*Nombre común + nombre común + adjetivo</i>	1,71
<i>*Nombre propio + nombre común</i>	1,60
<i>Nombre común + conjunción + nombre común</i>	1,52
<i>Nombre común + adjetivo + adjetivo</i>	0,89
<i>*Nombre común + adjetivo + nombre común</i>	0,49
<i>Otras</i>	10,25

Más del 75% de los enunciados son sintagmas nominales con la estructura propia del lenguaje natural (orden directo). Sin embargo, las categorías señaladas con un asterisco en la tabla suponen una alteración de esta estructura por un cambio en el orden de los términos o por la omisión de los nexos. Este cambio se adapta a la sintaxis *Encabezamiento + Subencabezamiento* de los índices de materia.

Civilización árabe historia Diccionario cine Elecciones Italia
Cine aspectos sociales Madrid historia Microbiología leche

Los enunciados que respetan el orden directo potencialmente pueden coincidir con un encabezamiento o con un subencabezamiento del índice. Ciñéndonos a las cuatro

categorías más importantes, los nombres comunes y los nombres propios son, en principio, la forma habitual de designar un concepto simple. En cuanto a las estructuras *nombre + adjetivo* y *nombre + preposición + nombre*, como ya se ha indicado en el primer capítulo, en los índices de materia se emplean para referirse conceptos más específicos que los denominados por un solo sustantivo pero no para expresar dos conceptos.

Los usuarios emplean correctamente el primer tipo, porque añadir un adjetivo a un nombre es la forma más normal de especificarlo. Sin embargo, cuando se trata de categorías en las que el modificador del sustantivo es un sintagma preposicional³, se observan tanto expresiones correctas como ejemplos en los que el nexo une términos que pertenecen a clases conceptualmente diferentes (algo más del 50% de los casos). Estos últimos ejemplos, si aparecieran en el índice lo harían en orden inverso y sin nexos (Encabezamiento – Subencabezamiento) con lo cual, la coincidencia sólo es posible si se eliminan las preposiciones y se utilizan técnicas de recuperación que permitan ignorar el orden de las palabras.

Un solo concepto

Comités de empresa

Medios de comunicación

Cuartetos de cuerda

Dos conceptos

Prostitución de inmigrantes

Historia del pensamiento político

Depuración de aguas

³ Nombre común + preposición + nombre común, Nombre común + preposición + artículo + nombre común, Nombre común + preposición + nombre común + adjetivo, Nombre común + preposición + nombre propio. En conjunto suman el 19,37% de los enunciados.

¿La distribución de las *características* es similar en todos *los grupos*? Comprobamos cierta similitud en cuanto al *número de palabras* ya que la media siempre está próxima al 2%, con un predominio claro de búsquedas con uno o dos términos (Tabla A.10 Anexo). Sin embargo, los porcentajes no son muy homogéneos entre los grupos y mucho menos cuando se trata de búsquedas con cuatro palabras o más.

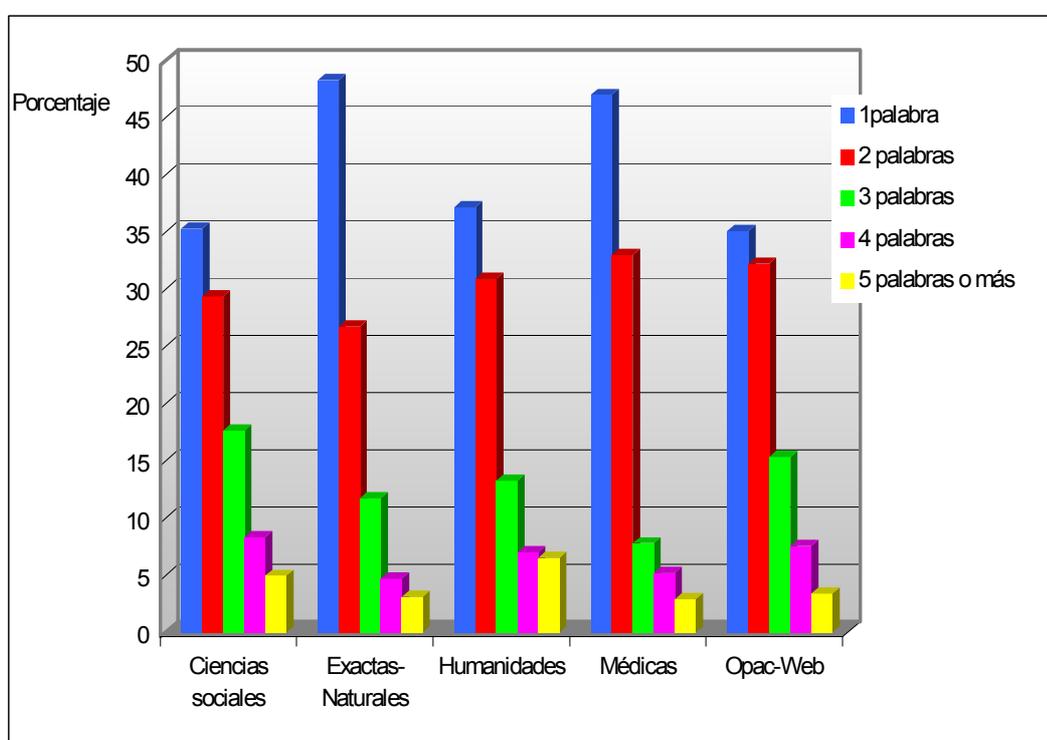


Fig.3.8. Número de palabras: distribución por áreas

El análisis de la varianza demuestra que las diferencias entre las áreas son significativas, aunque no en todos los casos (Tabla A.11 Anexo). No lo son entre las búsquedas realizadas desde las de Ciencias Sociales y Humanidades ($t=.686$), ni entre las de estas dos áreas y las del OPAC-Web ($t=.186$; $t=.351$). Tampoco lo son

las que existen entre los enunciados de Exactas-Naturales y los de Médicas ($t=.775$). Por tanto, esta característica de los enunciados parece diferenciar las búsquedas de las áreas científicas de las realizadas por los usuarios de “letras” y de los accesos vía Web. Una mirada al gráfico (Fig 3.8) confirma que las búsquedas de los “científicos” son más simples en el sentido de que las de una palabra o dos son muchos más numerosas que en los otros grupos.

Esta simplicidad se repite en relación con otras características. Por ejemplo, en la distribución de las *búsquedas por campos*, aunque existe un claro predominio de las búsquedas por términos de materia en todas ellas. A pesar de que esta superioridad es uniforme, la disparidad en la proporción de los otros tipos hace que existan diferencias significativas entre todos los grupos (Tabla A.13 Anexo).

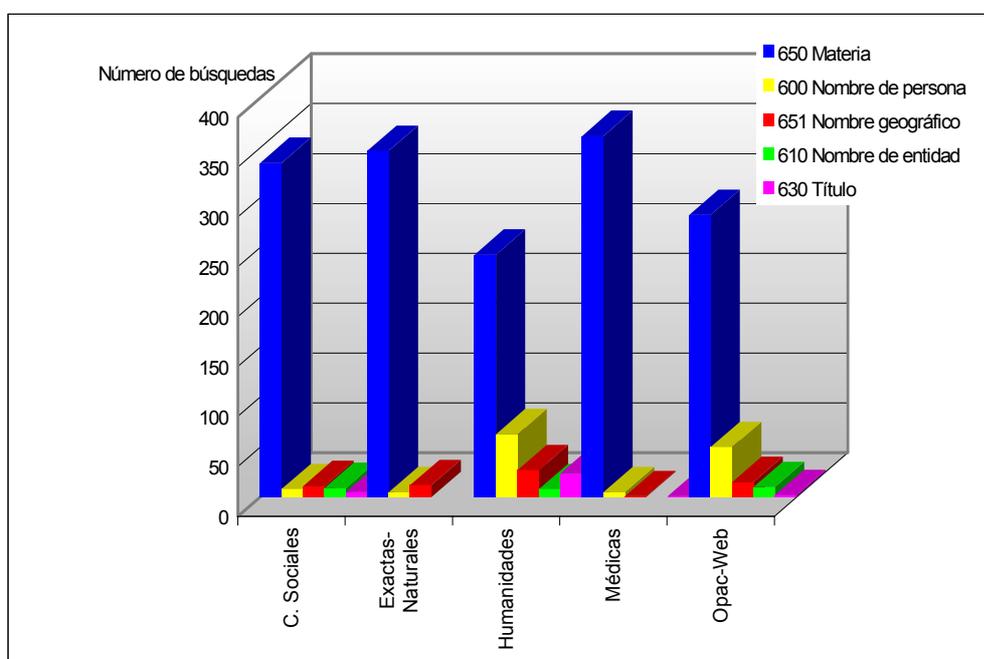


Fig 3.9 Tipos de búsqueda: distribución por áreas

Las áreas “de ciencias” vuelven a presentar cierta homogeneidad en esta característica y, a pesar de las diferencias, se muestran como las más similares. Son las que tienen búsquedas menos variadas porque las de materia representan más del 90% - en el caso de médicas el 94% (Tabla A.12 Anexo) – y como consecuencia casi no existen de las otras clases. En cualquiera de los dos grupos (Exactas-Naturales y Médicas) apenas aparecen los nombres de persona, no hay búsquedas por nombres de entidades y sólo una por título en el área médica. Probablemente la diferencia más destacable entre ellas es que en Ciencias Exactas-Naturales la búsqueda por nombres geográficos es la segunda categoría en importancia, aunque sólo representan un 3,12% del total.

El área con una distribución más peculiar y variada es la de Humanidades. En primer lugar, la proporción de búsquedas temáticas (63,12%) es sensiblemente menor que en las demás. Evidentemente, es así porque también es donde se plantean más búsquedas por nombres de personas (16,36%), lugares (7,01%) y títulos de obras literarias o artísticas (6,23%). Sin embargo, el número de enunciados de nombre de entidad es superior en las Ciencias Sociales o en el OPAC-Web. Las búsquedas de este último grupo parecen encontrarse en un punto intermedio entre las de “ciencias” y las de “letras”.

Como ya hemos visto, estas dos características - el número de palabras y el tipo de búsqueda - están relacionadas con la *estructura*. De nuevo, las diferencias que

pueden observarse en la tabla 3.4 son significativas entre todas las áreas aunque en todos los grupos predominan las búsquedas por nombre común o nombre común más adjetivo (Tabla A.14 Anexo). Sin embargo, las similitudes no van más allá porque incluso la distribución de estas dos categorías es poco homogénea. En la tabla se han destacado en negrita los datos sobre las categorías cuya frecuencia está por encima del 5% en cada grupo. Sólo las dos predominantes superan este porcentaje (equivalente a 20 búsquedas o más) en todas las áreas y otra más – nombres propios - en cuatro grupos de los cinco.

Tabla 3.4 Estructura: distribución por áreas

Estructura	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	Opac-Web
NC	116 30,13 %	170 44,16%	80 20,78 %	178 46,23 %	108 28,05 %
NC+Adj	85 22,08 %	70 18,18%	77 20 %	104 27,01 %	71 18,44 %
NC+Adj+Adj	3 0,78 %		4 1,04 %	10 2,60 %	3 0,78 %
NC+Adj+NC	2 0,52 %	1 0,26 %	2 0,52 %	1 0,26 %	2 0,52 %
NC+Conj+NC	6 1,56 %	4 1,04 %	7 1,82 %	6 1,56 %	5 1,30 %
NC+NC	12 3,12 %	20 5,19 %	6 1,56 %	16 4,16 %	15 3,90 %
NC+NC+Adj	8 2,08 %	2 0,52 %	7 1,82 %	1 0,26 %	6 1,56 %
NC+NP	5 1,30 %	3 0,78 %	10 2,60 %	2 0,52 %	8 2,08 %
NC+Prep+Art+NC	15 3,90 %	7 1,82 %	9 2,34 %	7 1,82 %	6 1,56 %
NC+Prep+NC	24 6,23 %	29 7,53 %	9 2,34 %	9 2,34 %	14 3,64 %
NC+Prep+NC+Adj	9 2,34 %	7 1,82 %	4 1,04 %	4 1,04 %	8 2,08 %
NC+Prep+NP	12 3,12 %	4 1,04 %	6 1,56 %	1 0,26 %	1 0,26 %
NP	30 7,79 %	21 5,45 %	89 23,12 %	5 1,30 %	61 15,84 %
NP+NC	6 1,56 %	2 0,52 %	4 1,04 %	1 0,26 %	10 2,60 %
Otras	35 9,09 %	24 6,23 %	51 13,25 %	24 6,23 %	42 10,91 %

La proporción más alta de búsquedas corresponde a la categoría nombre común, excepto en Humanidades donde son más numerosas las búsquedas por nombre propio. De esta forma, esta área vuelve a mostrar un comportamiento peculiar frente al resto de los grupos. El más similar en este caso es el OPAC-Web donde el porcentaje de las dos primeras categorías es más reducido, los nombres propios suponen más del 15% de las búsquedas y sólo estas tres categorías superan el 5%.

En el área de Ciencias Médicas, las dos categorías más importantes significan el 73,24% de las búsquedas y ninguna otra supera el 5%. En el extremo contrario, los enunciados de Exactas-Naturales tienen la distribución más variada de todos los grupos con cinco categorías por encima del umbral de las 20 ocurrencias, aunque las dos primeras (*Nombre común* y *Nombre común + Adjetivo*) significan más del 50%.

3.3 Similitud con el índice.

Casi un tercio de las expresiones de búsqueda utilizadas en el campo materias coinciden exactamente con las empleadas en el índice. Las referencias que reenvían a términos equivalentes (Véase) incrementan este porcentaje en un 3,33 %. Si se normaliza la sintaxis, ignorando el orden de las palabras y eliminando las vacías, la cifra asciende levemente alcanzando el 37,86%.

Tabla 3.5 Coincidencia de las expresiones con los términos del índice

		Sin correcciones	Corregido
Coincidencia exacta			
Orden directo		35,83	39,22
	Índice	32,50	35,35
	Véase	3,33	3,87
Cambio en el orden de las palabras		0,94	1,13
	Subcampos	0,86	1,04
	Palabras	0,08	0,08
Palabras significativas (sin vacías)		1,09	2,23
	Mismo orden	0,54	0,71
	Cambio de orden		
	Subcampos	0,45	1,42
	Palabras	0,10	0,10
Total		37,86	42,58
Coincidencia parcial			
Orden directo		10,89	11,07
	Índice	10,28	11,06
	Véase	0,61	0,01
Cambio en el orden de las palabras		6,56	7,50
	Subcampos	2,46	2,86
	Palabras	4,10	4,64
Palabras significativas (sin vacías)		2,16	2,38
	Mismo orden	0,36	0,36
	Cambio de orden		
	Subcampos	0,48	0,58
	Palabras	1,32	1,44
Total		19,61	20,97
Coincidencia palabras		29,59	25,93

Si a esta regularización le sumamos la corrección de los errores ortográficos y tipográficos, la normalización de las formas singular/plural, de las abreviadas y completas y de las fechas, se alcanza 42,58%. El factor que más peso tiene en este aumento es la corrección ortotipográfica – un 2,22% - seguida de la de las formas singular/plural (*Encuesta/ Encuestas, Loro/Loros, Empresa informativa/Empresas informativas*). Apenas se han encontrado diferencias que impidan la coincidencia

cuando se trata de fechas o abreviaturas entre otros motivos porque estos elementos aparecieron con poca frecuencia en la muestra analizada.

Tabla 3.6 Porcentaje corrección de errores sobre coincidencia

Coincidencia exacta	
Errores ortográficos/tipográficos	2,22
Singular/Plural	1,83
Busq. Singular/Encab plural	1,40
Busq. Plural/Encab singular	0,43
Abreviaturas/Fechas	0,45
Coincidencia parcial	
Errores ortográficos/tipográficos	1,04
Singular/Plural	0,48
Busq. Singular/Encab plural	0,40
Busq. Plural/Encab singular	0,08
Abreviaturas/Fechas	0,06

Además, la terminología utilizada por los usuarios coincide en un 19,61% con encabezamientos que contienen más términos (Coincidencia parcial). La mayor parte de estas coincidencias son enunciados que empiezan con los mismos términos que un encabezamiento más amplio (Orden directo) aunque la normalización de la sintaxis supone en este caso un considerable incremento de la equiparación (8,72%). Este aumento se debe fundamentalmente al cambio de orden de las palabras, porque al tener el encabezamiento más palabras que el enunciado se reduce la importancia de la eliminación o no de palabras vacías. Por el contrario, la corrección de errores eleva muy levemente el nivel de coincidencia en este grupo (1,36%), predominando de nuevo el factor de los ortotipográficos (Tabla 3.6).

Coincidencia parcial

Búsqueda

Talavera de la Reina
Cuerdas vocales
Infancia en Madrid
Trastornos de la alimentación
Segunda República historia

Encabezamiento

Platería – España – Talavera de la Reina
Perros – Cuerdas vocales - Cirugía
Infancia – España - Madrid
Ancianos – Alimentación - Trastornos
España – Historia – 1931-1939, Segunda República

Por último, un 29,59 % de las expresiones incluye algún término que aparece en uno de los encabezamientos del índice. Evidentemente, esta cifra se reduce si tenemos en cuenta las correcciones porque estas habrán significado una coincidencia con alguna categoría anterior. El resultado final es que sólo un 7,7% de los enunciados empleados por los usuarios contienen términos que nunca aparecen en el índice de materias y esta cifra se reduciría al 5,3% si se eliminara el grupo de errores que se han considerado.

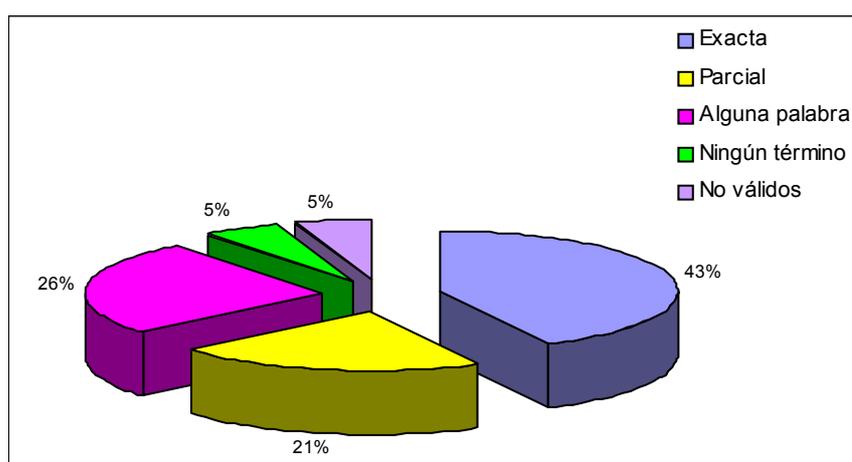


Fig. 3.10 Tipos de coincidencia: distribución total (con correcciones)

Como puede apreciarse en el gráfico, la coincidencia de alguna palabra significa el porcentaje más importante después de la exacta. A pesar de la importancia cuantitativa de esta categoría, los resultados parecen indicar que el nivel de similitud terminológica no es muy elevado, lo que significa que los encabezamientos recuperados probablemente no se ajusten demasiado al tema de búsqueda .

Tabla 3.7 Coincidencia parcial Índice/Enunciados

Coincidencia	Palabras en el índice		Palabras en un encabezamiento	
	% Sin correcciones	% Corregido	% Sin correcciones	% Corregido
1 palabra	6,95	5,12	21,70	18,75
2 palabras	14,84	13,32	6,80	6,19
3 palabras	6,72	6,41	1,00	0,99
4 palabras	0,97	0,97		
5 palabras	0,10	0,10		
Total	29,59	25,93	29,50	25,93

Tabla 3.8 Coincidencia palabras: palabras del enunciado en el mismo encabezamiento

	Nº palabras en un encabezamiento			
	0	1	2	3
2 palabras en el enunciado	10,09	54,67		
3 palabras en el enunciado	0,59	26,52	36,82	
4 palabras en el enunciado	2,35	12,29	34,00	46,47
5 palabras en el enunciado		3,57	16,04	44,87
6 palabras en el enunciado	2,35	1,57	9,32	1,31
7 palabras en el enunciado		0,55	1,68	7,35
8 palabras en el enunciado		0,14	0,97	
9 palabras en el enunciado			0,97	

La presencia de más de tres términos del enunciado en el índice es insignificante y si nos centramos en la presencia de las palabras en un mismo encabezamiento, que sería más significativo, comprobamos que la mayoría de las ocasiones sólo aparece una. Sin embargo, no se debería ignorar la importancia del 7,8% de los casos en que

coinciden dos o tres términos porque se trata fundamentalmente de enunciados de búsqueda de 3 a 5 palabras. Si tenemos en cuenta que en esta categoría se han eliminado las palabras vacías al contabilizar los términos comunes y no en el recuento de palabras por enunciado, las palabras que coinciden probablemente sean significativas. También puede ser interesante la similitud temática de los casos en que el enunciado coincide con términos de varios encabezamientos pero asignados todos a un mismo registro (3,5%).

Coincidencia de una palabra en el encabezamiento

Búsqueda	Encabezamientos
<i>Arte del Islam</i>	<i>Islam</i> ⁴ <i>Arte, Arte románico, Arte barroco, etc.</i> ⁵
<i>Identidad femenina</i>	<i>Identidad colectiva, Identidad cultural, etc.</i> ⁶ <i>Delincuencia femenina, Sexualidad femenina, etc.</i> ⁷
<i>Biodeterioro de metales</i>	<i>Granito – Biodeterioro</i> <i>Metales, Metales – Acabado, Metales – Análisis, etc.</i> ⁸
<i>Diseño de catalizadores</i>	<i>Diseño, Anuncios – Diseño, Calles - Diseño, etc.</i> ⁹ <i>Catalizadores, Catalizadores – Tesis, etc.</i> ¹⁰

Coincidencia de más de una palabra en un encabezamiento

Búsqueda	Encabezamiento
<i>Decadencia Dinastía Ming</i>	<i>China – Historia – 1368-1644, Dinastía Ming</i>
<i>Escritoras negras americanas</i>	<i>Escritoras negras – Estados Unidos – Historia y crítica</i>
<i>Trastorno de estrés postraumático</i>	<i>Estrés postraumático</i>
<i>Reinado de Isabel II</i>	<i>España – Historia – 1833-1868, Isabel II - Tesis</i>
<i>Manual de rehabilitación de la columna vertebral</i>	<i>Columna vertebral – Enfermedades - Rehabilitación</i>

⁴ El término Islam aparece en 286 encabezamientos.

⁵ El término arte aparece en más de 5000 encabezamientos.

⁶ El término identidad aparece en 90 encabezamientos.

⁷ El término femenina aparece en 90 encabezamientos.

⁸ El término metales aparece en 238 encabezamientos.

⁹ El término diseño aparece en 640 encabezamientos.

¹⁰ El término catalizadores aparece en 7 encabezamientos.

Coincidencia de las palabras en un mismo registro

Búsqueda	Registro
<i>Financiación autonómica</i>	<i>Comunidades autónomas – Financiación Hacienda autonómica – España</i>
<i>Biblia milagros</i>	<i>Biblia. N.T. – Crítica e interpretación Milagros – Edad antigua</i>
<i>Antibióticos microorganismos resistencia</i>	<i>Antibióticos Microorganismos – Resistencia a los medicamentos</i>
<i>Aprendizaje por refuerzo</i>	<i>Aprendizaje Refuerzo (Psicología)</i>
<i>Legislación y ética profesional</i>	<i>Psiquiatría – Legislación Psiquiatras – Ética profesional</i>

Pese a la importancia cuantitativa de la coincidencia de palabras, las categorías que suponen el mayor porcentaje de la equiparación son la *coincidencia exacta* y *parcial*. Además, son las que permitirían recuperar encabezamientos más parecidos al enunciado de búsqueda y, por tanto, potencialmente más relevantes para resolver la necesidad de información.

La relación entre las características de las expresiones de búsqueda que hemos visto en el apartado anterior y los dos tipos de coincidencia es significativa en todos los casos (con y sin correcciones). La media del *número de palabras* de los enunciados que coinciden exacta o parcialmente con los términos del índice es ligeramente inferior a la media general debido al peso que tienen en la equiparación las búsquedas de un término (Tabla 3.9). También destaca en ambos casos el porcentaje de *búsquedas temáticas*, aunque la presencia de los nombres de persona experimenta un importante incremento en la coincidencia parcial. En cuanto a la *estructura*, los resultados están en estrecha relación con las dos características

anteriores: son más similares al índice de materias las estructuras más simples y los nombres propios coinciden mayoritariamente de forma parcial con los encabezamientos.

Tabla 3.9 Coincidencia exacta y parcial/características: palabras y tipo

	%Exacta	%Parcial		%Exacta	%Parcial
1 palabra	51,47	49,27	600 Nombre de persona	1,66	28,31
2 palabras	31,35	29,77	610 Nombre de entidad	1,85	2,19
3 palabras	11,68	14,17	630 Título	1,70	1,42
4 palabras	3,63	4,58	650 Materia	90,04	61,18
5 palabras	1,06%	1,52	651 Nombre geográfico	4,74	6,90
6 palabras	0,66	0,68			
7 palabras	0,16				
Media	1,66	1,77			

Tabla 3.10 Coincidencia exacta y parcial/características: estructura

	% Exacta	% Parcial
Nombre común	46,74	26,45
Nombre común + Adjetivo	24,54	10,59
Nombre común + Adjetivo + Adjetivo	0,23	
Nombre común + Adjetivo + Nombre común	0,57	0,28
Nombre común + Conjunción + Nombre común	0,95	0,84
Nombre común + Nombre común	3,12	2,81
Nombre común + Nombre común + Adjetivo	0,80	2,46
Nombre común + Nombre propio	1,54	1,40
Nombre común + Preposición + Artículo + Nombre común	2,23	1,67
Nombre común + Preposición + Nombre común	4,62	3,49
Nombre común + Preposición + Nombre común + Adjetivo	0,40	
Nombre común + Preposición + Nombre propio	1,13	3,02
Nombre propio	6,69	35,31
Nombre propio + Nombre común	1,46	2,89
Otras	4,98	8,79

Estas distribuciones se explican también porque el porcentaje de enunciados con estas características es tan importante que era previsible que predominaran también en cualquier tipo de equiparación. Si analizamos la proporción de cada tipo de

expresión que coincide total o parcialmente con el índice conseguimos una perspectiva probablemente más significativa, aunque en este caso resulte bastante similar. En los siguientes gráficos se han representado los resultados desde este punto de vista.

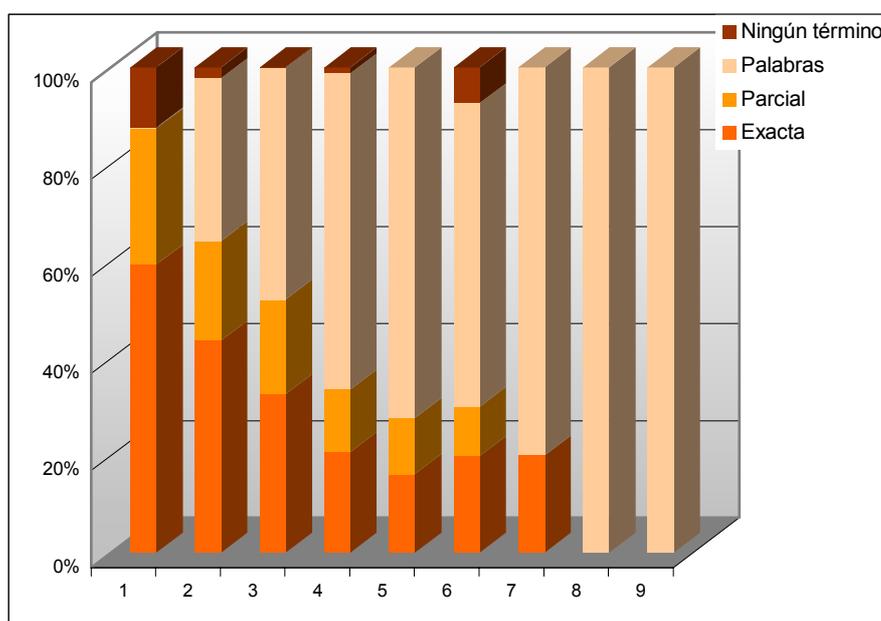


Fig. 3.11 Número de palabras/Tipo de coincidencia

Respecto al *número de palabras* se confirman los resultados anteriores ya que las posibilidades de coincidir exacta o parcialmente con el índice disminuyen a medida que se incrementa el número de palabras del enunciado de búsqueda. Esta afirmación sólo sería matizable para las búsquedas que tienen entre 4 y 7 palabras en las que se observan oscilaciones en la tendencia.

Sin embargo, sí se aprecian algunos matices nuevos cuando la característica es el *tipo de búsqueda*. La coincidencia exacta de los enunciados que empiezan por o son un nombre de lugar supera a la coincidencia de los enunciados temáticos, aunque en ambos casos el porcentaje ronda el 50% (Tabla A.19 Anexo). Son bastante inferiores las proporciones de los mismos tipos de equiparación en el caso de los títulos o nombres de entidades y muy inferior la posibilidad de que una búsqueda por nombre propio de persona coincida exactamente con el encabezamiento.

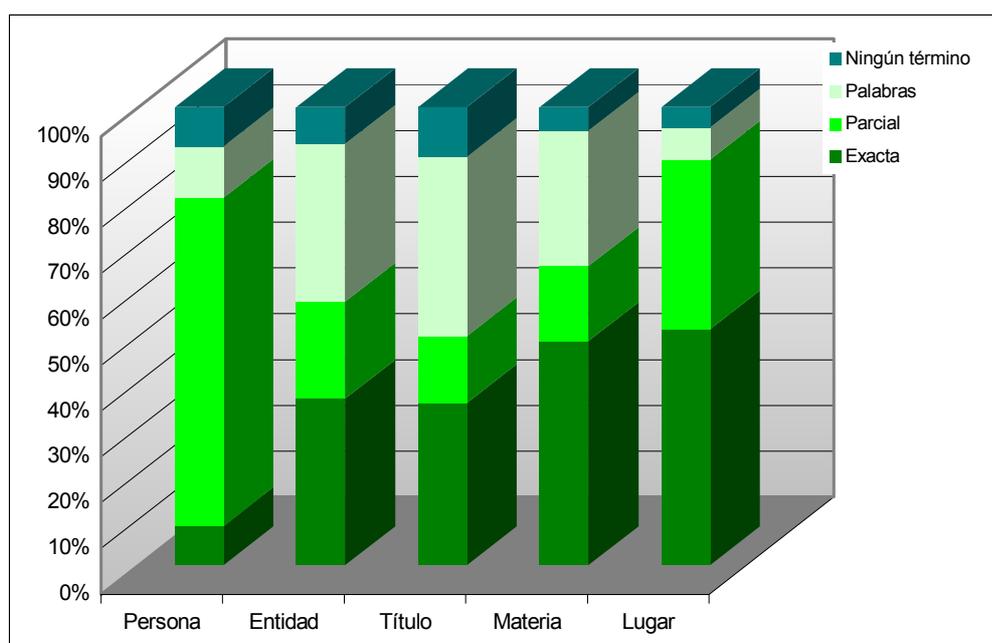


Fig. 3.12 Tipo de búsqueda/Tipo de coincidencia

Esta última circunstancia se explica por varios factores. El primero es que, como ya hemos visto, los usuarios frecuentemente utilizan sólo una parte del nombre, el apellido, cuando buscan información sobre una persona. Además, en el índice de materias del catálogo CISNE los nombres de personas van acompañados

habitualmente por las fechas de nacimiento y defunción, elementos que los usuarios no incluyen en las búsquedas, lo cual impide que la coincidencia sea exacta. Por último, la mayoría de los encabezamientos de nombres de personas incluyen algún subencabezamiento que especifica el tipo de estudio o de documento de que se trata, especificación que normalmente tampoco aparece en los enunciados. Esta situación se confirma por la importancia de la proporción de coincidencia parcial de esta categoría (71, 68%) que casi duplica a la que la sigue en importancia, el nombre de lugar (37%).

Búsqueda	Encabezamiento
<i>Hemingway</i>	<i>Hemingway, Esnest, 1899-1961</i>
<i>Gombrowicz Witold</i>	<i>Gombrowicz Witold, 1904-1969 – Diarios personales</i>
<i>Feuerbach Ludwig</i>	<i>Feuerbach, Ludwig, 1804 1872 - Critica e interpretación</i>
<i>Fernando VI</i>	<i>Fernando VI, Rey de España, 1713-1759</i>

Tampoco varían sustancialmente los resultados que se obtienen sobre la estructura de esta manera. Los enunciados con más posibilidades de coincidir exactamente son los que tienen una estructura sintáctica sencilla y aparecen en orden directo: *Nombre común*, *Nombre común + Adjetivo*, *Nombre común + Preposición + Nombre común*, aunque sólo las dos primeras superan el 50% (Tabla A.20 Anexo). Además, se corrobora la importancia de la coincidencia parcial para los enunciados que contienen *Nombres propios*.

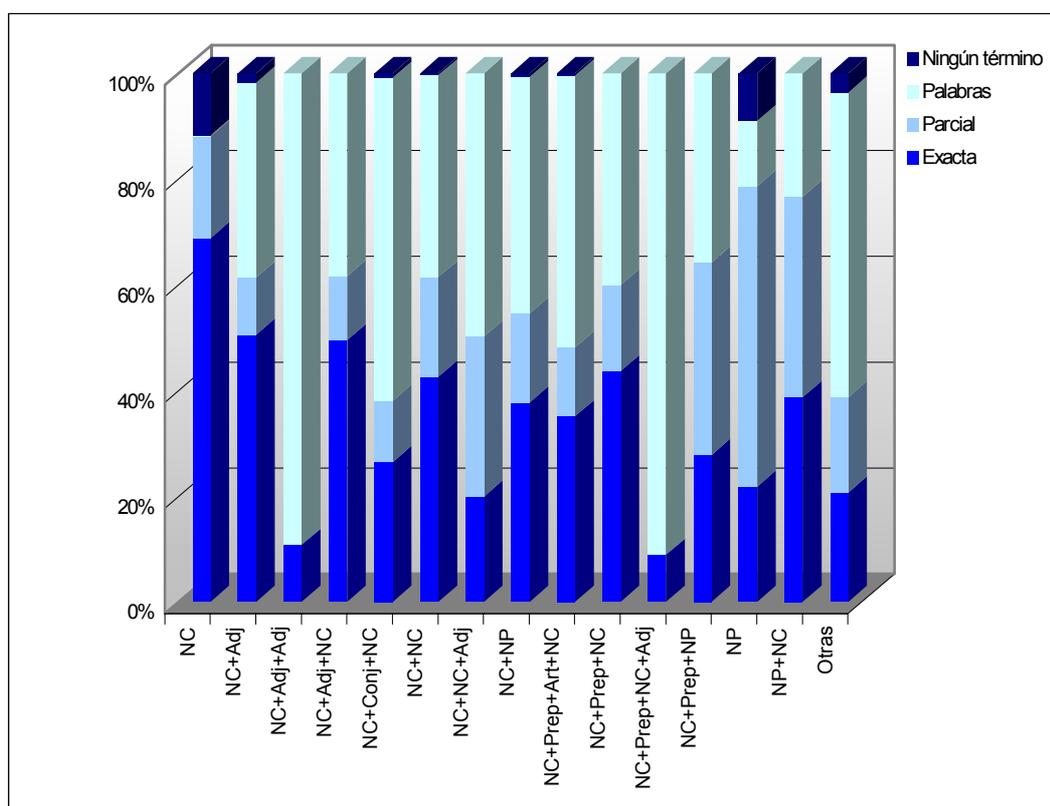


Fig. 3.13 Estructura/Tipo de coincidencia

Sin embargo, desde esta perspectiva se aprecia que también es importante el nivel de coincidencia de aquellas categorías que suponen una alteración del orden habitual de la estructura de los sintagmas nominales para adaptarse a la sintaxis de los encabezamientos de materia: *Nombre común + Adjetivo + Nombre común* (49,48%), *Nombre común + Nombre común* (42,56%), *Nombre propio + Nombre común* (38,74%) y *Nombre común + Nombre propio* (37,70%).

Por último queda valorar la coincidencia de los enunciados de búsqueda de las cuatro áreas temáticas y del OPAC-Web.

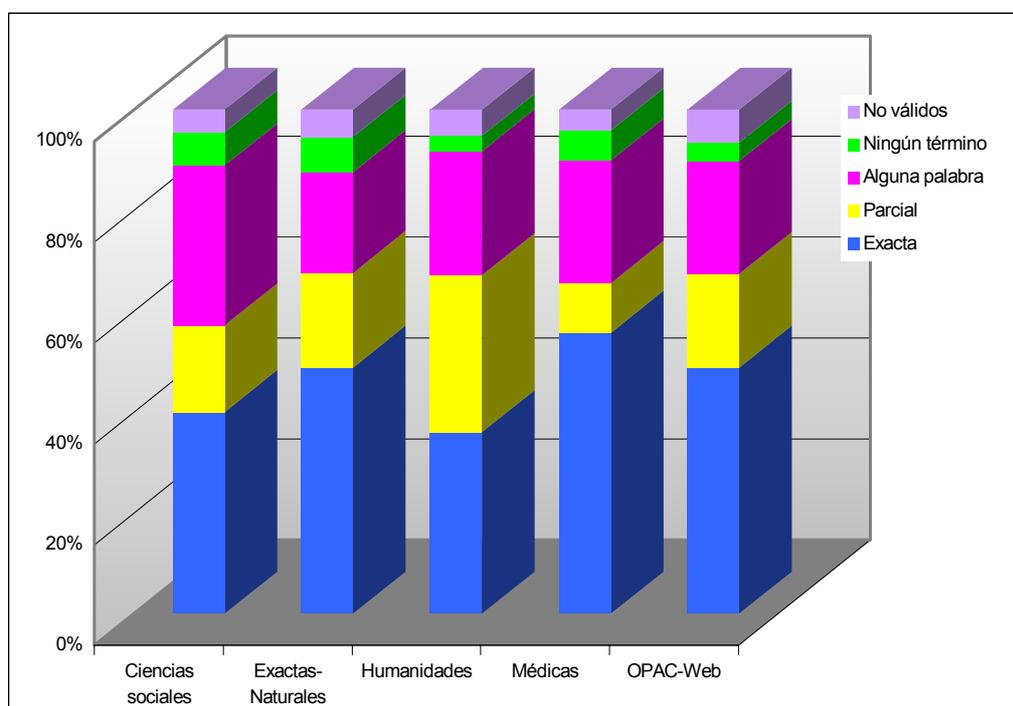


Fig. 3.14 Áreas temáticas: tipo de coincidencia

Aunque la diferencia entre los grupos es significativa tanto para la coincidencia exacta como para la parcial, el nivel que se alcanza sumando ambas categorías es muy similar en todos ellos – en torno al 66% –, excepto en las Ciencias Sociales que es inferior –57%– (Tablas A.21-A.25 Anexo). Esta cifra incide directamente en los resultados generales porque las búsquedas de esta área temática tienen un peso importante ya que representan el 40% del total.

Analizando por grupos la relación entre las características de las búsquedas y la coincidencia también se observan algunas variaciones. La relación es significativa en todos los casos cuando se trata de la coincidencia exacta ($p=.000$, Tabla A.29

Anexo). Sin embargo, en el caso de la parcial, que en los resultados generales también era significativa, presenta algunas variaciones entre las áreas. Sí se cumple esta relación en el caso de la estructura, pero no es significativo el tipo de búsqueda en las Ciencias Sociales, ni el número de palabras para esta misma área y para las “de ciencias” (Tabla A.30 Anexo).

Finalmente, hemos comparado la coincidencia exacta y parcial de los grupos para comprobar si había similitudes. En cuanto a la coincidencia exacta hay diferencias que permiten diferenciar las búsquedas de los grupos “de letras”¹¹ (Ciencias Sociales y Humanidades) de los “de ciencias” (Exactas-Naturales y Médicas¹²). El OPAC-Web sólo presenta similitudes con la equiparación de las Ciencias Exactas-Naturales.

Tabla 3.11 Coincidencia exacta por áreas. Prueba de Chi-cuadrado

	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	OPAC Web
Ciencias sociales					
Exactas-Naturales	.007				
Humanidades	.334	.000			
Médicas	.000	.097	.000		
Opac Web	.017	.829	.001	.041	

Estas diferencias no persisten en la coincidencia parcial, respecto a la que sólo dos grupos presentan un comportamiento “significativamente diferente”: las Humanidades y las Ciencias Médicas.

¹¹ Ciencias Sociales y Humanidades

¹² Exactas-Naturales y Médicas

Tabla 3.12 Coincidencia parcial por áreas. Prueba de Chi-cuadrado

	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	OPAC Web
Ciencias sociales					
Exactas-Naturales	.851				
Humanidades	.000	.000			
Médicas	.001	.002	.000		
Opac Web	.714	.518	.001	.000	

Si lo relacionamos con las características vemos que, efectivamente, se trata de las dos áreas que tenían enunciados con perfiles más claros y definidos. Por una parte, los enunciados de Humanidades incluían más palabras, tipos de búsqueda más variados, más nombres de personas y estructuras más complejas. En el extremo contrario estaban las búsquedas “Médicas” con menos palabras por enunciado, temáticas casi en su totalidad y muy simples estructuralmente.

4. Discusión

En esta sección se comentan los aspectos más destacables de los resultados que se han obtenido en esta investigación y se comparan con los de otros trabajos que analizan algún aspecto similar. En muchas ocasiones la metodología y las categorías utilizadas son tan diferentes que la comparación ha resultado complicada. Sin embargo, no se ha renunciado a ofrecer los datos más significativos porque pueden ayudar a contextualizar tanto los resultados generales como las diferencias que se han apreciado entre las áreas temáticas.

4.1 Distribución general.

4.1.1 Aspectos formales.

Los resultados generales del análisis de los enunciados de búsqueda del OPAC del catálogo CISNE no difieren sustancialmente de los obtenidos en los estudios que sirvieron para establecer las hipótesis de trabajo. A continuación, y siguiendo el mismo orden que en la exposición de los resultados, se comparan los datos y se comentan las diferencias metodológicas.

Como ya se ha indicado en el apartado de errores formales, las búsquedas no válidas y los errores ortotipográficos suman un 11,1% de los enunciados analizados. Se trata de expresiones que nunca coincidirían con los términos del índice, si no se corrigen o se establece un sistema de equivalencias.

La mitad de este porcentaje son búsquedas no válidas (5,2%). Aunque la categorización que se ha hecho de ellas varía de unos trabajos a otros, este resultado es sensiblemente inferior al 8,8% obtenido por Drabenstott y Vizine-Goetz (1994) o al 10,4% de Drabenstott y Weller (1996c). Las diferencias parecen equilibrarse si tenemos en cuenta que en estas dos investigaciones se incluyeron entre los enunciados “no válidos” aquellos que se identificaron como búsquedas por ejemplar conocido, entendiendo por tales las formadas por nombres de persona, autor-título, título y nombres de entidades (siempre que no fueran seguidas de un subencabezamiento temático)

Sin embargo, este equilibrio vuelve a romperse al comprobar que la categoría más importante de búsquedas no válidas en el OPAC del catálogo CISNE, lengua extranjera (51%), no se utilizaba en estos dos trabajos. La importancia que adquieren este tipo de enunciados parece deberse tanto al uso de neologismos en el lenguaje científico-técnico como a los accesos externos a través de Internet, puesto que los dos grupos en los que el porcentaje es mayor son OPAC-Web y Ciencias Exactas-Naturales.

La proporción de errores ortotipográficos encontrados y el predominio de los de omisión son resultados que coinciden, en mayor o menor medida, con los de estudios anteriores para las búsquedas en el campo materias (Tabla 4.1). La distribución de los fallos por campos de búsqueda también es similar a la obtenida, por ejemplo, por Drabenstott y Vizine-Goetz (1994).

Tabla 4.1. Errores ortotipográficos: datos comparados

Estudio	Campo de búsqueda	% Errores tipográficos
Connaway et al. (1995)	Todos	15,14% ¹
Drabenstott y Vizine-Goetz (1994)	Materias	3,54%
Drabensttot y Weller (1996c)	Materias	5,7%
Ferl y Millsap (1996)	Todos	11,05%
Jones (1986)	Materias	5,3%
Lester (1989)	Materias	5%
Walker y Jones (1987)	Materias	5,9%
CISNE	Materias	5,9%

En la Tabla 4.1 puede comprobarse que los porcentajes de este tipo de errores son menores cuando sólo se analizan enunciados del campo materias. Si tenemos en cuenta que dentro de este campo los mayores porcentajes de errores tipográficos corresponden a las búsquedas por títulos, nombres de entidades y personas (Tabla 3.1), no resulta extraño que en aquellos trabajos que tienen en cuenta todos los campos de búsqueda el número de errores sea superior. Según nuestros datos esto se explica porque son búsquedas más largas – títulos y entidades habitualmente contienen dos o más palabras –, en las que es más frecuente el error de omisión, y

en el caso de los nombres propios porque muchas veces se utiliza una grafía incorrecta.

Lo que no deja de resultar sorprendente es que todavía no se incorporen de manera regular a los catálogos en línea mecanismos que permitan detectar y corregir este tipo de errores. Especialmente cuando se trata de herramientas de uso frecuente en sistemas de recuperación como los buscadores de Internet y cuando los experimentos realizados en OPAC's demuestran que mejoran sensiblemente los resultados de búsqueda (Walker y Jones 1987). En el caso de CISNE la corrección de este grupo de errores supondría un incremento del 3,6% de la coincidencia entre enunciados e índice. De ellos el 2,22% serían enunciados que coincidirían exactamente y en el 1,04% de los casos la coincidencia sería parcial. El resto de las búsquedas que tenían errores ortotipográficos contenían términos que no estaban en el índice a pesar de su corrección.

En cuanto a las características de los enunciados, los usuarios del OPAC de la Biblioteca de la Universidad Complutense de Madrid utilizan mayoritariamente expresiones simples que reproducen el orden de los sintagmas en el lenguaje natural. Además, las búsquedas se refieren a lugares y temas antes que a personas, entidades u obras originales identificadas por su título.

¹ Porcentaje calculado a partir de los datos que aparecen en el artículo (Connaway et al. 1995)

De la simplicidad de los enunciados da buena idea la media del número de palabras que es 2,1. Los resultados de las investigaciones que han trabajado con usuarios finales son bastante similares: ligeramente inferior para búsquedas en otros OPAC's – 1,8 (Drabenstott y Vizine-Goetz 1994) – y exactamente igual para el buscador Excite (Jansen, Spink y Saracevic 2000). Estos resultados se desvían significativamente de los obtenidos en trabajos sobre sistemas de recuperación de información comerciales para acceder a bases de datos bibliográficas, en los que el número medio de términos por búsqueda se mueve en un rango de 7-15 (Bates, Wide y Siegfried 1993; Fenichel 1981; Hsieh-Yee 1993; Spink y Saracevic 1997).

Hay que tener en cuenta que estos últimos estudios difieren no sólo en las herramientas de búsqueda sino también en el tipo de usuario objeto de análisis, ya que se trataría de especialistas en la materia o en búsqueda de información. Además, la metodología es tan distinta que es difícil comparar los resultados. Por ejemplo, no coinciden las definiciones de término de búsqueda – en unos cada palabra y en otros cada expresión – ni las de enunciado de búsqueda – en unos cada secuencia seguida por un intro y en otros una estrategia completa. Sin embargo, lo que sí parece confirmarse es que los usuarios con experiencia o conocimiento temático utilizan más términos de búsqueda y más específicos que los usuarios finales, quienes emplean menos términos y más genéricos.

Respecto a los tipos de búsqueda por materias, las búsquedas temáticas representan una proporción superior a la de otros estudios y, consecuentemente, el

resto de las categorías tiene proporciones inferiores (Tabla 4.2). Hay que recordar que además Drabenstott y Vizine-Goetz (1994) y Drabenstott (1996a) sólo consideraron búsquedas temáticas las expresiones formadas por nombres de persona, autor-título, título y nombres de entidades cuando incluían algún subencabezamiento. Esta diferencia metodológica hace que en comparación la proporción de búsquedas temáticas en CISNE sea todavía más importante.

Tabla 4.2. Campos de búsqueda: datos comparados

	Materia	Personas	Entidades	Lugar	Título
Drabestott y Vizine-Goetz (1994) ²	69,5 %	10,9%	2,6%	4,7%	3,3%
Bates, Wilde y Siegfried (1993) ³	51,3%	37,7%	—	6,3%	4,7 %
Lester (1989)	70%	—	—	—	—
CISNE	83,33%	5,57%	3,27%	1,74%	0,85%

El porcentaje de búsquedas por un tema también es superior al obtenido por Lester (1989) y mucho más alto que el de Bates, Wilde y Siegfried (1993). En el último caso puede explicarse porque en su investigación sólo se analizan enunciados de especialistas en Humanidades, área donde las búsquedas por nombres de personas o lugares habitualmente son mucho más numerosas.

Los datos relacionados con la estructura de los enunciados de búsqueda no son comparables con otros trabajos parecidos, por ejemplo el de Bates, Wilde y

² Los datos que se manejan en Drabenstott (1996) son los mismos. De hecho, los porcentajes sólo varían por redondeo: Materias 69% Nombres de personas 11% Geográfico 5%, Corporativo 3% Otros 12%.

³ Datos recalculados para facilitar la comparación.

Siegfried (1993), puesto que no coinciden las categorías empleadas en el análisis. Sintetizando los resultados obtenidos en CISNE podemos decir que aunque la estructura sintáctica de las expresiones que se han analizado es variada, predominan los *sustantivos* y la construcción *sustantivo + adjetivo*. La sintaxis de los enunciados coincide la mayoría de las veces con la de las expresiones en lenguaje natural, aunque en algunas ocasiones se utilizan estructuras que suponen una alteración de este orden: *nombre propio+nombre común, nombre común+nombre común+adjetivo, nombre común+nombre propio*. Esto se puede entender como una adaptación de los usuarios a las características de los índices, a la sintaxis de los encabezamientos, que puede ser producto de un aprendizaje en interacciones previas o una simple mimetización de los encabezamientos que se han visualizado.

4.1.2 Coincidencia con el índice de encabezamientos de materia.

Tampoco resulta fácil comparar los resultados que valoran la coincidencia entre los enunciados y el índice porque no sólo se han utilizado variables diferentes, sino también técnicas distintas para realizar las búsquedas. Con objeto de simplificar la presentación de los datos se han elaborado dos tablas. La primera (Tabla 4.3) contiene porcentajes generales de coincidencia, indicando el tipo de búsqueda que se ha empleado en el estudio. La segunda (Tabla 4.4) recoge los resultados de los dos trabajos que utilizan variables más similares a las empleadas en esta investigación y que permiten comprobar diferentes grados de coincidencia sin que

estén vinculados a un sistema de búsqueda concreto. Aun teniendo en cuenta estas indicaciones, los resultados deben mirarse con precaución.

Tabla 4.3 Coincidencia enunciado/índice: datos comparados

Estudio	Búsqueda	Coincidencia
Carlyle (1989)	Normalizado	60%+14%
Drabenstott y Vizine-Goetz (1994) ⁴	Normalizado	33%+16,1%
Drabenstott (1996a)	Normalizado	43,4%+13,1% ⁵
Hunter (1991)	Frase	38%
Larson (1986)	Palabras clave	51%
Lester (1989)	Frase	44%
Markey ⁶ (1984)	Palabras clave	50%
Tagliacozzo y Kochen (1995)	Palabras clave	57%
Wilkes y Nelson (1995)	Frase	41%
CISNE	Normalizado	42,6%+20,97 ⁷

Respecto a los tipos de búsqueda, se han estudiado catálogos que ejecutan búsquedas por palabras clave (Kern-Simirenko 1983; Larson 1986) y otros que utilizan índices “de frase” (Hunter 1991; Lester 1989; Wilkes y Nelson 1995). La utilización del primer tipo de búsqueda elimina las dificultades relacionadas con la sintaxis de los encabezamientos – no importa el orden de las palabras –, reduciendo el problema de la coincidencia a la terminología. En consecuencia, el nivel de

⁴ Datos recalculados para facilitar la comparación

⁵ Estos porcentajes están calculados sobre el total de las búsquedas aunque en el artículo aparecen también las proporciones sólo sobre enunciados válidos (Drabenstott 1996a)

⁶ Modificada por Carlyle (1989)

⁷ Se han utilizado las cifras relativas a los enunciados “normalizados” para facilitar la comparación. Aunque se comentará más adelante, recordamos que los niveles de coincidencia sin corrección y sin ningún tipo de normalización son: coincidencia exacta 35,83% y coincidencia parcial 10,89%.

similitud es mayor: si se respeta la sintaxis está entorno al 40% mientras que la utilización de palabras clave significa un incremento de aproximadamente un 10%.

Por último, en los estudios de Carlyle (1989), Drabenstott y Vizine-Goetz (1994), Drabenstott (1996a) y en CISNE, la equiparación se ha realizado después de una “normalización de los enunciados” que incluye tanto la eliminación de palabras vacías y cambios en el orden de las palabras como la corrección de errores ortotipográficos, la eliminación de diferencias singular/plural, la normalización de las abreviaturas y de la forma de las fechas. Para estos cuatro trabajos se presentan las cifras de coincidencia exacta y coincidencia parcial que pueden verse con más detalle en la Tabla 4.4.

Tabla 4.4 Tipo de coincidencia enunciado/índice: datos comparados

	Carlyle 89	Drabenstott y Vizine-Goetz 94	Drabestott 96a	CISNE
Coincidencia exacta	60%	33%	43,4%	42,58%
Coincidencia parcial	14%	16,2%	13,2%	20,97%
Algún término en el índice	21%	—	15,1%	25,93%
Términos que no aparecen en el índice	5%	—	—	5,3% (+5,2% no válidos)

Las cifras que aparecen en los trabajos de Drabenstott y Vizine-Goetz (1994) y de Drabenstott (1996a) no permiten hacerse una idea clara del incremento de la coincidencia que supone la “normalización” en las diferentes categorías. Los datos son más diáfanos en el caso del estudio de Carlyle (1989) que aparecen en la Tabla

4.5 junto a los resultados obtenidos del análisis de las búsquedas en el catálogo de la Universidad Complutense.

Tabla 4.5 Normalización por categorías de coincidencia: datos comparados

	Carlyle 89	CISNE
Coincidencia exacta		
Sin “normalizar”	47 %	35,83%
Cambio en el orden de las palabras y eliminación de nexos.	4,5%	2,25%
Corrección errores tipográficos, diferencias singular/plural, abreviaturas, etc.	8,5%	4,5%
Total “normalización”	13%	6,75%
Total	60%	42,58%
Coincidencia parcial		
Sin “normalizar”	10%	10,89%
Cambio en el orden de las palabras y eliminación de nexos.	Sin especificar	8,72%
Corrección errores tipográficos, diferencias singular/plural, abreviaturas, etc.	Sin especificar	1,36%
Total “normalización”	4%	10,08%
Total	14%	20,97%

Como puede apreciarse, aunque las mayores proporciones en cada categoría se consiguen sin ningún tipo de modificación del enunciado, la “normalización” contribuye a mejorar la equiparación en un porcentaje sustancial: sumando ambos grupos la coincidencia se incrementa alrededor del 17%. En el caso de CISNE, la corrección de errores y otras diferencias formales supone casi un 6%, con más peso en el porcentaje de coincidencia exacta (aunque sensiblemente inferior al del estudio de Carlyle).

La normalización del orden de las palabras y los nexos en el caso de la coincidencia exacta no es muy importante y afecta casi exclusivamente al orden de los subcampos (Tabla 3.5). No sucede lo mismo con la coincidencia parcial que se incrementa en un 8,72% si se buscan las palabras en cualquier orden y se eliminan los nexos o, en otros términos, si se realiza una búsqueda por palabras clave eliminando las palabras vacías.

En la Tabla 4.4 puede comprobarse que los porcentajes obtenidos en las diferentes categorías de equiparación son bastante dispares. En el análisis realizado por Carlyle (1989) el nivel de coincidencia exacta es muy elevado mientras que en el de Drabenstott y Vizine-Goetz (1994) es muy escaso. En el trabajo de Carlyle (1989) el reducido tamaño de la muestra (200 enunciados) desgraciadamente limita la generalización de los resultados⁸, mientras que en el de Drabenstott y Vizine-Goetz se han tenido que recalcular los datos a partir de una presentación de los resultados por tipos de búsqueda. Teniendo en cuenta estos factores, parece que la información más válida para la comparación es la del estudio de Drabenstott (1996a). Cotejándolo con este trabajo, en CISNE se ha alcanzado un nivel ligeramente inferior de coincidencia exacta mientras que los niveles de coincidencia parcial y de enunciados que contienen algún término que está en el índice son sensiblemente superiores.

⁸ Es una lástima que sea así porque en otros aspectos metodológicos, como la definición de las variables - el trabajo es el más preciso de los consultados.

Si miramos los resultados de menor a mayor nivel de coincidencia, en primer lugar encontramos el 25,93 % que corresponde a la categoría “algún término en el encabezamiento” con lo que la probabilidad de que la información recuperada sea relevante disminuye notablemente. Como ya se ha indicado en el apartado de resultados, quizá la información más pertinente de este grupo correspondería al 7,18% de coincidencia de dos o tres términos en el mismo encabezamiento o el 3,5% de términos en encabezamientos asignados a un mismo registro.

Los niveles de la categoría “coincidencia parcial” son bastante superiores a los de otras investigaciones lo que parece indicar que se emplean expresiones que son menos específicas que los encabezamientos que aparecen en el índice. Este menor nivel de especificidad no hay que interpretarlo en un sentido jerárquico sino que debe entenderse como “especificidad horizontal”, relacionada con el grado de precoordinación de los índices (Lancaster 1996; Slype 1991; Soergel 1994). Es decir, los encabezamientos del índice son más concretos porque se utilizan más subdivisiones, más términos, para delimitarlos.

Teniendo en cuenta que la media de palabras empleada en la búsqueda es similar o superior a la de otros estudios sobre catálogos en línea en el ámbito anglosajón, el alto porcentaje de coincidencia parcial parece indicar que, efectivamente en CISNE se emplean más subdivisiones. Este dato puede estar relacionado con la tendencia a la simplificación en el uso de las subdivisiones propuesta para la LCSH, en la que se llegó a plantear la separación de los subencabezamientos de lugar y forma a

campos diferentes del formato MARC (Chan 1992; Conway 1992). También es cierto que, por ejemplo, la política de indización de la National Library of Medicine dispone de hecho que estos dos tipos de subdivisiones se consignen en campos separados que se deben combinar en el momento de la búsqueda (Application of MeSH for Medical Catalogers). Sin embargo, con los datos que se han manejado en esta investigación sólo es posible apuntar esta posibilidad. Para confirmarla y generalizarla a los índices en castellano sería necesario analizar una muestra de los encabezamientos empleados como puntos de acceso.

Por último, la “coincidencia exacta” es muy inferior a la de Carlyle (1989) y ligeramente menor que en Drabenstott (1996a) aunque se mantiene en la línea de esta y de otras investigaciones (Tabla 4.4). También hay similitudes en cuanto a las características de los enunciados que tienen más posibilidades de coincidir con el índice. En síntesis son búsquedas más sencillas, con menor número de palabras y que tratan sobre lugares o temas. La media del número de palabras de las búsquedas que coinciden es menor que el de las que no coincide⁹ y, obviamente, las estructuras sintácticas también son más simples, especialmente para la coincidencia exacta.

En cuanto al tipo de enunciado, en CISNE las búsquedas que alcanzan mayor nivel de coincidencia exacta con el índice son las de nombres de lugar (51,53%) y las temáticas (48,93%). La importancia de esta última categoría es similar a los resultados de otras investigaciones (Drabenstott 1996a; Lester 1989) en las que

también predominó la coincidencia de las búsquedas temáticas. Sin embargo, es muy superior el nivel de expresiones que empiezan por un nombre geográfico y que coinciden exactamente. El 50% de las veces esta coincidencia exacta se consigue cuando se ha empleado un término geográfico sin subdivisiones, por lo tanto con puntos de acceso muy genéricos (*Tajo, Chile*). El 50% restante son enunciados en los que se añaden especificaciones utilizando correctamente la sintaxis de los encabezamientos de materia que en este caso difiere del uso natural de la lengua (*Alemania historia, México elecciones*). Ya hemos indicado como posibles explicaciones de esta situación la adaptación de los usuarios al sistema o la reproducción de encabezamientos que se han visualizado.

Quizá uno de los datos que más llama la atención es el escaso porcentaje de coincidencia exacta de los nombres de personas (8,52%) frente al elevado nivel que se consigue en esta categoría cuando la equiparación es parcial (71,68%). En este caso influyen dos factores: el primero es que en el 50% de estas búsquedas los usuarios emplean sólo un apellido, y el segundo que los encabezamientos de personas del índice incluyen más información que la que indican los usuarios (subencabezamientos y/o fechas de nacimiento y defunción).

Finalmente, habría que señalar que las relaciones de equivalencia reflejadas en el catálogo mediante la referencia Véase sólo incrementan la similitud en un 4,55%¹⁰.

⁹ Frente a 2,1 de media general, la de los enunciados que coinciden exactamente es 1,66 y aquellos que coinciden parcialmente tienen de media 1,77 palabras.

¹⁰ 3,87% en la coincidencia exacta y 0,68% en la parcial.

La cifra es inferior a la alcanzada en el estudio de Lester (1989) que era del 10,6% o en el de Markey (1984)¹¹ que supuso un 7%. Por tanto, la diferencia indica una menor utilización de la estructura semántica de los encabezamientos para mejorar el acceso por materias en CISNE. La infrautilización se produce a pesar de las facilidades tecnológicas desarrolladas en el periodo de tiempo que separa los trabajos. En consecuencia, parece confirmarse la opinión de quienes critican la deficiente estructura sindética de los catálogos en línea (Cochrane 2000; Greenberg 1997) y defienden la necesidad de mejorar el uso del catálogo de autoridades para la recuperación de información (Tartaglia 2003; Tillet 2000)

4. 2 Diferencias entre áreas temáticas.

Como se indicó en las hipótesis de trabajo, las investigaciones empíricas han revelado que estudiantes e investigadores de distintas áreas de conocimiento buscan información de distinta manera. De nuevo hay que insistir en que las diferencias metodológicas hacen difíciles las comparaciones. En este caso son más significativas que en el anterior lo que ha obligado a hacer sólo valoraciones de conjunto y no comparaciones detalladas.

¹¹ Citado por Carlyle (1989)

Estas diferencias son mayores en primer lugar porque la mayoría de los estudios se refieren a usuarios de bases de datos bibliográficas que forman un conjunto de población – investigadores y especialistas en la materia – menos heterogéneo que el de los OPAC's. En segundo lugar, porque se analiza el comportamiento de búsqueda en general, no sólo los enunciados. Este objetivo implica la utilización de métodos de corte cualitativo como la observación, protocolos verbales y teoría fundamentada.

Teniendo en cuenta estas apreciaciones, los resultados de las investigaciones previas apuntan en el sentido de diferentes “hábitos documentales” de los usuarios de distintas especialidades científicas a la hora de resolver sus necesidades de información. De hecho, varios estudios sugieren tendencias más “conservadoras” en el acceso y uso de la información entre los especialistas en Humanidades y Ciencias Sociales. El uso que hacen estos científicos de la biblioteca es más tradicional: trabajan más aislados y sus fondos particulares a menudo solapan las colecciones de la biblioteca; encuentran más fácilmente información en libros y revistas y, por tanto, valoran en menor medida otros servicios de información (Cool 1993; Stone 1982; Swigger 1992; Tibbo 1994).

Si nos centramos en el proceso de búsqueda de información, el experimento realizado por Borgman (1986) demostró que los especialistas en áreas científicas y técnicas obtienen mejores resultados de los actuales sistemas de recuperación que los de Ciencias Sociales y Humanidades. La razón fundamental de esta diferencia es que

el modelo que predomina en estos sistemas, las búsquedas analíticas basadas en la lógica booleana, requiere poner en funcionamiento habilidades de razonamiento lógico y simbólico que se adaptan mejor a la formación recibida en las áreas científico-técnicas (Borgman 1989).

Las diferencias pueden matizarse si se miran los resultados de algunos trabajos. Por ejemplo, Ellis, Cox y Hall (1993) no encontraron diferencias significativas entre el comportamiento de búsqueda de investigadores en Física y en Ciencias Sociales. Aunque no se tuvieron en cuenta los resultados obtenidos en la búsqueda, las actividades realizadas durante el proceso no indicaron conductas sustancialmente distintas.

No sucede lo mismo con los especialistas en Humanidades quienes, como señala Shaw (1995), se distinguen tanto de los “científicos” como de los usuarios de Ciencias Sociales. Stone (1982) ya había apuntado su tendencia a delegar las búsquedas bibliográficas, su necesidad de explorar la información y su empleo de puntos de acceso específicos – fechas y lugares –, su mayor utilización de monografías y la tendencia a hacer búsquedas retrospectivas. Los estudios de Wiberly (1983), Wiberly y Jones (1988), Wiberly (1991) Cool (1993) o Bates, Wilde y Siegfried (1995) ratifican estas afirmaciones. En una revisión sobre la bibliografía dedicada a la materia, Tibbo (1991) sintetizaba este comportamiento indicando que la búsqueda tiende a ser más intuitiva, menos premeditada, y con un objetivo menos definido que suele cambiar durante el proceso. Además, los

principales puntos de acceso son bastante diferentes: por ejemplo, se utilizan con frecuencia nombres de personas, lugares y acontecimientos.

En el trabajo de Bates, Wilde y Siegfried (1995) se analiza específicamente el vocabulario y las características de los enunciados de búsqueda de los “humanistas”, comparándolos con la información de Saracevic et al. (1988) para Ciencias Naturales y Ciencias Sociales. Las conclusiones sobre los puntos de acceso utilizados son similares a las de Tibbo (1991): se emplean muchos más nombres propios con los que se plantean búsquedas temáticas relativas a personas, lugares, acontecimientos y títulos de obras (literarias o plásticas). La especificidad de estas búsquedas por nombres propios contrasta con el carácter genérico de aquellas que se refieren a materias propiamente dichas en la que se utiliza con frecuencia nombres de disciplinas.

¿En qué medida coinciden estos resultados con los que se han obtenido del análisis de los enunciados utilizados en CISNE?

Respecto a las características de las expresiones de búsqueda, se han apreciado diferencias entre las áreas aunque en muchos casos no hayan sido significativas. En cuanto al número de palabras se perfilan dos grupos: uno formado por las búsquedas de Ciencias Sociales, Humanidades y OPAC-Web y otro de “ciencias” formado por las de Ciencias Exactas-Naturales y Ciencias Médicas. Los datos indican que los enunciados del segundo grupo tienen menos palabras que los del primero.

El tipo de búsqueda no permite diferenciar grupos afines puesto que hay diferencias significativas entre todas las áreas. Pese a estas diferencias, las búsquedas de “ciencias” resultan más “parecidas” porque la proporción de enunciados temáticos es superior y porque casi no se utilizan nombres propios. En el extremo opuesto las búsquedas más variadas son las de Humanidades donde la reducción de enunciados temáticos implica la utilización más nombres propios de personas, lugares, títulos y entidades.

Respecto a la estructura también hay diferencias significativas entre todas las áreas pese al predominio de los *nombres comunes* y de la agrupación *nombre común + adjetivo*. De nuevo la excepción a esta última afirmación la constituye el área de Humanidades, porque es la única donde el mayor porcentaje de búsquedas corresponde a la categoría *nombre propio*. En este grupo se incluyen tanto los nombres de personas como los de entidades o lugares, con lo que se reafirman los resultados obtenidos respecto al tipo de búsqueda.

Teniendo en cuenta que se ha comprobado que existe relación entre las características y la coincidencia con el índice y dada la aparente desigualdad entre las características de los enunciados de las diferentes áreas temáticas cabía esperar que también se produjera desigualdad entre ellas en cuanto a la equiparación. La Tabla 4.6 sintetiza la información de las tablas A.21-A.25 del Anexo, datos que en el apartado de resultados se presentaron de forma gráfica (Fig. 3.14).

Tabla 4.6 Tipo de coincidencia enunciado/índice por áreas

	Coincidencia exacta	Coincidencia parcial	Algún término
Ciencias sociales	39,74%	13,2%	31,57%
Exactas-Naturales	48,83%	18,70%	20%
Humanidades	36,10%	31,17%	24,64%
Médicas	55,84%	9,87%	24,42%
OPAC-Web	48,31%	18,44%	22,18%

Como puede observarse, los niveles de similitud varían bastante en función de que la equiparación sea exacta o parcial. En el primer caso los niveles de significación alcanzados en la prueba de Chi-cuadrado confirman la división “letras” (Ciencias Sociales y Humanidades) y “ciencias” (Exactas-Naturales y Médicas). Sin embargo, el nivel de coincidencia exacta más parecido es el de OPAC-Web y Exactas-Naturales. Respecto a la coincidencia parcial solo son significativamente diferentes los grupos de Humanidades y Ciencias Médicas.

Por tanto, los resultados confirman las peculiaridades de las búsquedas de los “humanistas” ya planteadas en otros estudios: más búsquedas por nombres propios (lugares, personas y entidades) y títulos, estructuralmente más variadas y con más palabras (Bates, Wilde y Siegfried 1995; Cool 1993; Tibbo 1991; Wiberly 1991; Wiberly y Jones 1988). Un factor determinante para las diferencias en la equiparación puede ser el hecho de que la proporción de nombres de personas sea muy superior a la del resto de los grupos, dado que es un tipo de enunciado con un comportamiento muy

diferente a los demás ya que apenas coincide exactamente con el índice pero sí tiene un alto nivel de coincidencia parcial.

Estos rasgos contrastan con los de la otra área que presenta un comportamiento diferente: Ciencias Médicas. En este caso, la coincidencia es mayoritariamente exacta y los enunciados se caracterizan tener pocas palabras, simplicidad de estructuras y referirse a búsquedas casi exclusivamente temáticas. Estas características son habituales en las búsquedas “de ciencias” según el análisis que hacen Bates, Wilde y Siegfried (1993) de los datos del estudio de Saracevic y Kantor (1988) sobre las búsquedas realizadas en la National Science Foundation.

5. Conclusiones

En este apartado se exponen las conclusiones de la investigación relacionadas tanto con los aspectos metodológicos como con los resultados obtenidos. Así mismo, se recogen y comentan algunas de las alternativas propuestas para solucionar los problemas que plantea la equiparación de los enunciados de búsqueda con el índice de materias. Finalmente, se sugieren algunos temas de investigación para posibles estudios cuyos resultados podrían complementar y matizar este trabajo.

5.1 Conclusiones generales.

Quizá uno de los aspectos más interesantes y más complejos de cualquier investigación sea el relativo a las decisiones que afectan a la metodología. Por esta razón esta sección de conclusiones se inicia con una recapitulación de algunas de las reflexiones que sobre materiales y método se han planteado a lo largo del trabajo.

El análisis de las transacciones ha resultado un método válido para el tipo de estudio que se ha realizado. Es cierto, como ya se ha señalado, que no proporciona información sobre el comportamiento de búsqueda, ni sobre la relación entre la necesidad de información y la expresión utilizada, ni sobre la relevancia de los resultados, etc. Indiscutiblemente, si se quiere realizar un estudio de este tipo habría que recurrir a fuentes de datos y métodos de análisis de corte más cualitativo (observación, análisis de protocolos verbales, etc.). Sin embargo, el análisis de las transacciones permite conocer con detalle las expresiones que han utilizado los usuarios sin que ningún tipo de factor externo condicione la búsqueda. De hecho, las variables utilizadas han permitido obtener una detallada descripción de las características de las búsquedas y de los niveles de similitud con el índice.

También en relación con la metodología habría que insistir en los problemas que plantea la variedad de técnicas de recogida de datos, variables y métodos de análisis que se han utilizado. Esta heterogeneidad afecta no sólo a aquellas investigaciones que analizan comportamientos de búsqueda de tipo genérico sino a las que han analizado los mismos aspectos que se han estudiado en este trabajo. La diversidad no sólo dificulta la comparación de los resultados, además pone de manifiesto la falta de un paradigma de referencia que sirva de modelo para la investigación empírica. Debilidad que, como señalan Yuan y Meadow (1999), puede considerarse un síntoma de la juventud de una disciplina incipiente, cuyas bases epistemológicas son consideradas todavía por algunos como poco sólidas.

Respecto a los aspectos formales de los enunciados de búsqueda, los resultados confirman las hipótesis de trabajo de las que se partía, lo que indica un alto nivel de coincidencia con los datos de estudios anteriores para la lengua inglesa.

En el apartado de enunciados no válidos se repiten las categorías de otros trabajos a las que hay que añadir las búsquedas en lengua extranjera, predominantemente inglés. Este último grupo parece vinculado a las búsquedas a través de Internet y a los neologismos científico-técnicos.

También se han encontrado errores ortotipográficos (aproximadamente un 5%) sin que se hayan apreciado grandes diferencias en las proporciones respecto a estudios anteriores (Drabenstott y Weller 1996c; Jones 1986; Lester 1989; Walker y Jones 1987). La corrección de estos fallos formales, habitual en otros sistemas de recuperación de información, podría mejorar los resultados de la equiparación hasta un 3,6% en el caso del catálogo CISNE.

Los resultados sobre las características de los enunciados confirman que los usuarios de los catálogos en línea utilizan pocas palabras para expresar las materias sobre las que necesitan información. Aunque la estructura sintáctica de las expresiones es muy variada, predominan las más simples y en orden directo (*nombre* o *nombre + adjetivo*). Por último, el tipo de búsqueda más frecuente es sobre un tema (83%) frente a los escasos porcentajes de las que se refieren a personas, lugares, entidades o títulos de obras literarias y cinematográficas.

Estos resultados parecen confirmar el carácter genérico de las búsquedas de los usuarios de catálogos en línea (Drabenstott 1996a; Hancock-Beaulieu 1987; Jones, Cunnigham y McNab 1998; Pulis y Ludy 1988; Walker y Jones 1987). Es cierto que si nos atenemos a una especificidad puramente jerárquica y sólo se valoran los datos cuantitativos, se puede considerar una afirmación matizable, porque en ocasiones una sola palabra pueda ser más específica que dos o tres (por ejemplo, desarrollo cognitivo y sincretismo).

Sin embargo, hay que tener en cuenta que una forma “natural” de concretar el significado de un sustantivo es añadirle otras palabras para formar un sintagma nominal. Además habría que recordar que la especificidad, cuando se utilizan encabezamientos de materia, se alcanza también gracias a la precoordinación de varios términos (especificidad horizontal). Por tanto, en función tanto de la estructura como del número de palabras de los enunciados, sí se puede hablar de búsquedas genéricas, al menos bastante menos concretas que las realizadas por especialistas en la materia o en recuperación de información. Estos enunciados permiten al usuario acceder a la información que está almacenada en el catálogo, pero probablemente sean necesarias posteriores especificaciones o limitaciones para reducir el conjunto de resultados obtenido.

Respecto a la coincidencia, los niveles de coincidencia exacta teniendo en cuenta la sintaxis y sin ningún tipo de corrección de errores ortotipográficos, de las

diferencias singular/plural o de las abreviaturas, son algo inferiores a los de otros estudios (35, 83%). Sin embargo, si se realiza esta “normalización” se alcanza un 42,6% que es una proporción similar a la de otras investigaciones. Los enunciados que más posibilidades tienen de coincidir en esta categoría se refieren a lugares o temas, tienen menos palabras que la media y estructuras sintácticas más simples. En consecuencia, coinciden con encabezamientos genéricos y con bastantes posibilidades de aparecer en el índice también con subdivisiones. Por tanto se trata de enunciados que, si se utilizan búsquedas por palabras clave o con truncamiento a la derecha, probablemente recuperen muchos registros en catálogos de gran tamaño.

El nivel de coincidencia parcial y de aparición de alguna palabra en el índice es muy importante: sumando las dos categorías alcanza un 47%. En el caso de la categoría coincidencia parcial (20,97%) se trata de enunciados que, aunque son más genéricos que los puntos de acceso del índice, podrían recuperar encabezamientos o registros potencialmente útiles, especialmente para iniciar una búsqueda exploratoria. Esta “aproximación” cuando coinciden términos y sintaxis podría conseguirse haciendo una búsqueda con truncamiento a la derecha que recuperaría aquellos encabezamientos que “empiezan como” la búsqueda. Cuando la coincidencia es sólo terminológica habría que ejecutar la búsqueda por palabras clave, con lo que la ordenación de los encabezamientos recuperados resultaría más complicada (Wool 2000).

En cuanto a los enunciados que sólo coinciden en algún término con el índice (25,93%), evidentemente hay muchas menos posibilidades de que exista parecido semántico, disminuyendo a medida que se reduce el número de términos en que coinciden. Sin embargo, su recuperación también podría contribuir a que el usuario continuara explorando el sistema en lugar de abandonar la búsqueda porque no ha obtenido ningún resultado.

Por lo que acabamos de exponer hasta aquí, resulta evidente que existe una relación entre las características de los enunciados y la equiparación: tienen más posibilidades de coincidir las búsquedas temáticas, con pocos términos y estructuras simples. Uno de los grupos que presenta un comportamiento más particular son las búsquedas por nombres de personas:

- ✓ La mayor parte de las veces sólo se producirá una coincidencia parcial, tanto porque se utiliza un sólo un apellido en los enunciados como porque los índices contienen especificaciones que no son empleadas por los usuarios: fechas de nacimiento y defunción, subdivisiones de materia.

Búsqueda: Proust

Proust, Louis Joseph, 1754-1826

Proust, Marcel, 1871-1922

Proust, Marcel, 1871-1922 – Biografías

Proust, Marcel, 1871-1922 – Correspondencia

Proust, Marcel, 1871-1922 – Crítica e interpretación

Proust, Marcel, 1871-1922 – Temas y motivos

Balzac, Honoré de, 1799-1850 – Influencia – Proust, Marcel, 1871-1922

Flaubert, Gustave, 1821-1880 – Influencia – Proust, Marcel, 1871-1922

- ✓ Cuando se emplea también el nombre de la persona, muchas veces es el primer elemento del enunciado. Como consecuencia la coincidencia sólo se conseguirá si se busca por palabras clave porque es necesario que no importe el orden de los términos dentro de un mismo subcampo.

Búsqueda: Charlotte Brontë

Brontë, Charlotte, 1816-1855 -- Biografías

Brontë, Charlotte, 1816-1855 -- Crítica e interpretación

Brontë, Charlotte, 1816-1855. Jane Eyre -- Crítica e interpretación

Brontë, Charlotte, 1816-1855. Shirley -- Crítica e interpretación

Brontë, Charlotte, 1816-1855. Villette -- Crítica e interpretación

A la vista de los resultados se puede afirmar que aproximadamente un 90% de los enunciados empleados por los usuarios del catálogo CISNE incluyen algún término que aparece en el índice de encabezamientos de materia¹. En otras palabras, si se emplearan técnicas de equiparación parcial² con algún tipo de ordenación por relevancia de los encabezamientos o registros recuperados, sólo un 10,5% de las búsquedas no ofrecerían información al usuario. Por tanto, para mejorar las búsquedas por materias parece necesario mejorar las técnicas de recuperación por equiparación.

Pero los mismos datos indican que tanto o más importante es la necesidad de facilitar herramientas que ayuden a explorar el sistema y a especificar las búsquedas. En la interacción con el sistema, como en cualquier otro proceso de

¹ Sin normalizar sería el 87,06%

comunicación, el nivel de eficacia dependerá de la terminología y la sintaxis que son las que van a permitir emitir un mensaje, expresar un contenido sin el que el mensaje no existiría. Si el usuario no utiliza correctamente o no conoce el “lenguaje” del sistema, difícilmente podrá comunicarle su necesidad de información. Evidentemente, una de las soluciones es que el usuario aprenda este lenguaje pero la realidad es que cada día son más numerosos los usuarios finales que consultan catálogos sin ninguna formación previa. En esta situación, la balanza parece inclinarse hacia una segunda opción que consiste en ayudar al usuario y enseñarle durante la interacción. No se trata sólo de simplificar las búsquedas analíticas sino también de mostrarle lo que puede encontrar allí, cómo está organizado y ayudarle a encontrar las palabras o expresiones precisas para que puedan entenderse.

Concretamente, si tenemos en cuenta los resultados que se han obtenido en la equiparación entre enunciados de búsqueda y puntos de acceso temáticos en CISNE dos razones fundamentales justifican la visualización de los índices:

- ✓ Muchos usuarios inician sus búsquedas con términos que coinciden exactamente con los del índice pero que son muy genéricos, probablemente más que los temas que realmente les interesan. La consulta de los índices de materia puede ayudarles a concretar la búsqueda.

² Para ver la relación entre las técnicas de recuperación y los tipos de índices se puede consultar la

- ✓ La importancia cuantitativa de las dos categorías que indican que se recuperan encabezamientos más o menos “parecidos” a la búsqueda pero no exactamente iguales es muy elevada. En ambos casos se incrementa la cantidad de información recuperada (tasa de llamada) aunque se reduce la posibilidad de que sea relevante (tasa de precisión). La visualización de los índices en este caso puede invitar a iniciar una búsqueda exploratoria durante la cual es posible concretar el enunciado, completarlo o encontrar términos de búsqueda alternativos.

Finalmente, sintetizando las diferencias que se han encontrado entre las áreas temáticas se puede decir que algunas características de los enunciados y el comportamiento en la equiparación permiten establecer ciertas similitudes entre las áreas de “ciencias” y las de “letras”. Sin embargo, sólo dos grupos presentan perfiles claramente diferentes: Humanidades y Ciencias Médicas. En cuanto a los rasgos de las búsquedas realizadas desde el OPAC-Web, no constituyen un grupo homogéneo claramente distinto de los demás.

La peculiaridad de las búsquedas de Ciencias Médicas radica en su sencillez: son búsquedas temáticas casi en su totalidad, con menos palabras que en otras áreas y estructura sintáctica más sencilla. Con estos rasgos se produce un alto nivel de coincidencia exacta, mientras que se reduce notablemente la parcial y se mantiene la aparición de algún término en el índice dentro de la media. Por tanto, los

enunciados o bien son similares en el nivel de especificidad con los índices (muy genéricos) o contienen términos que aparecen dispersos en varios puntos de acceso.

En el extremo opuesto, las expresiones de Humanidades se caracterizan por ser más largas y variadas. La variedad se refiere fundamentalmente a los tipos de búsqueda ya que en esta área son mucho más numerosas las búsquedas por nombres propios. La abundancia de enunciados que se refieren a nombres de lugares, entidades o personas coincide con el perfil de las búsquedas de los “humanistas” que se presenta en otras investigaciones (Bates, Wilde y Siegfried 1995; Tibbo 1991, 1994). Respecto a la coincidencia, la exacta es la más baja de todas las áreas y la parcial la más alta con diferencia (casi dobla a las otras áreas). El elevado nivel de coincidencia parcial está condicionado en gran medida por las búsquedas de nombres propios de persona en las que, como ya se ha indicado, los usuarios incluyen menos información que la que aparece en los índices.

De estas diferencias se desprenden algunas consecuencias que consideramos interesante señalar:

- ✓ Algunas características de los enunciados tienen mayor incidencia en el nivel de coincidencia que el área temática a la que correspondan las búsquedas. Concretamente el tipo de búsqueda, sobre todo nombres propios de personas, y el número de palabras parecen factores decisivos. De hecho, los “árboles de búsqueda” propuestos por Drabentstott y Vizine-Goetz (1994) o Drabentstott

(1996a) separan en primer lugar las búsquedas por nombres de personas y después diferencian entre las de un término y las de más de uno.

- ✓ Teniendo en cuenta que, como señala Soergel (1994), los niveles de recuperación –coincidencia en este caso – dependen directamente de cómo está indizada la información, las diferencias entre estas dos áreas parecen indicar diferencias no sólo en los enunciados sino también en los índices. Si las búsquedas de Ciencias Médicas son más cortas y simples y hay mayor nivel de coincidencia exacta mientras que las de Humanidades son más largas y hay mayor índice de coincidencia parcial, con muchas precauciones se puede considerar que los índices de Humanidades son más específicos en el sentido de que se emplean más términos en cada punto de acceso.

5.2 Propuestas de mejora.

En el estado de la cuestión se han expuesto las propuestas más relevantes para mejorar las búsquedas por materias en los catálogos en línea. Resulta sorprendente que después de más de dos décadas de investigación muchos de los problemas se hayan paliado en sistemas experimentales e incluso en algunos sistemas de bases de datos o buscadores de Internet y sin embargo estas soluciones no se hayan

implementado todavía en los OPAC's comerciales. A continuación se plantean aquellas que parecen más necesarias a la luz de los resultados de esta investigación.

En primer lugar, habría que insistir en la necesidad de incorporar a los catálogos en línea algún medio para corregir los errores ortográficos y tipográficos, lo que potencialmente supondría mejorar los resultados en torno al 4%. Se trata de implementar algoritmos similares a los descritos por Walker y Jones (1987) que permitan la detección de errores y su corrección automática o semiautomática (dando al usuario la posibilidad de corregirlos él mismo). Una posible alternativa sería que, cuando no se recuperan registros, el sistema al menos mostrara al usuario el término de búsqueda que ha tecleado pidiéndole que lo revise para comprobar si está bien escrito.

Respecto a otras diferencias formales como puede ser la falta de coincidencia en el número de las palabras se han propuesto diferentes soluciones:

- ✓ La utilización de algoritmos de lematización para eliminar sufijos, entre ellos los plurales. Sin embargo, las características de las búsquedas en los OPAC's no parecen hacerlos recomendables porque al tratarse de búsquedas muy genéricas el uso de estos algoritmos incrementaría excesivamente el volumen de encabezamientos o registros recuperados.

- ✓ La forma más simple de corregir esta diferencia es utilizar sistemáticamente el truncamiento a la derecha para que se recuperen los términos en plural aunque se hayan escrito en singular. Drabenstott y Vizine-Goetz (1994) proponen que además siempre que el sistema detecte una terminación plural (-s, -es) la elimine y se realice la búsqueda con un truncamiento a la derecha. Se trata de un sistema económico pero que produce tanto o más ruido que el anterior.

- ✓ Otra alternativa es incorporar al catálogo de autoridades las formas singular y plural del mismo término como formas equivalentes y realizar la consulta simultáneamente con las dos siempre que sólo una de ellas sea un término admitido (Carlyle 1989; Drabenstott y Vizine-Goetz 1994).

La creación de registros de autoridad también parece una buena solución para eliminar otras diferencias formales como las distintas maneras de expresar las fechas o las siglas y sus formas desarrolladas (incluyendo las diferentes opciones como términos equivalentes). Sin embargo, para lo que se ha considerado primordial un buen catálogo de autoridades es para incrementar el vocabulario de entrada incluyendo todos términos alternativos para designar un mismo concepto (Tartaglia 2003; Tillet 2000). En el estudio realizado por Wilkes y Nelson (1995) se comprobó que el 73,9% de los enunciados que no recuperaban registros tenían entradas alternativas en el catálogo de autoridades. Si hubiera existido un enlace

con los registros de autoridad, no se habría producido este error. En el caso de las búsquedas que se han analizado del catálogo CISNE sólo un 4,5% de las coincidencias se debe a la presencia de relaciones de equivalencia.

Aunque la importancia del control de autoridades para la recuperación de información es cada día más evidente (Tartaglia 2003; Tillet 2003) parece que la información de los registros de autoridad muchas veces va más dirigida a mejorar la calidad de los procesos técnicos bibliotecarios que a facilitar las búsquedas. De hecho, es poco frecuente que estos datos se exploten totalmente en los sistemas de búsqueda. La mayoría de las veces sólo se informa al usuario de su existencia con una terminología poco familiar para él. Como señalan Cochrane (2000) o Greenberg (1997), pese a las oportunidades que ofrecen las nuevas tecnologías, la estructura sindética de los catálogos en línea continúa siendo deficiente, en ocasiones incluso peor que la de los catálogos manuales.

Otro bloque de mejoras es el relacionado con las técnicas de recuperación de información utilizadas para equiparar búsquedas e índices. Viendo los niveles de coincidencia alcanzados en CISNE sí parece adecuada una combinación de las técnicas de recuperación por equiparación exacta utilizadas habitualmente en los catálogos en línea³, a las que habría que añadir técnicas de equiparación parcial.

³ Búsqueda sobre el campo completo con una lectura “secuencial” izquierda-derecha y búsqueda por palabras clave considerando implícito el operador booleano Y.

En los trabajos de Dabenstott y Vizine-Goetz (1994) y Drabenstott (1996a) se propone un árbol de búsqueda que combina estas técnicas para mejorar la recuperación. En primer lugar diferencia las búsquedas por nombres propios de personas de las temáticas. A continuación se establece una secuencia de tipos de equiparación que se justifica porque va de mayor a menor posibilidad de parecido semántico en función del parecido terminológico y sintáctico: coincidencia exacta, coincidencia exacta con truncamiento a la derecha, palabras clave en el encabezamiento y palabras clave en el registro.

Los resultados obtenidos en CISNE invitan a simplificar los árboles de búsqueda construyendo una secuencia con sólo dos pasos: cuando no se produce la equiparación en el primero, pasaría a ejecutarse la segunda opción de búsqueda. Esta segunda posibilidad incorpora el uso de técnicas de equiparación parcial que todavía no se utilizan habitualmente en los catálogos en línea (aunque sí en bases de datos y buscadores de Internet). El orden de las técnicas también está en relación con la potencial equivalencia entre parecido formal y semántico. Respecto a la conveniencia de separar los nombres de persona, es una opción que no deja de tener ventajas aunque consideramos que la secuencia propuesta también es válida para las características de este tipo de enunciados.

El cuadro de la Figura 5.1 sintetiza la propuesta indicando la técnica de búsqueda, el tipo de índice necesario, el tipo de visualización más indicado y el porcentaje

aproximado de enunciados que coincidirían con el índice en cada caso según los resultados obtenidos en este trabajo.

Técnica	Índices	Visualización	Porcentaje
1. – Equiparación exacta Lectura secuencial (izquierda-derecha) Truncamiento a la derecha	Encabezamientos con permutación de subcampos	Índice alfabético de materias Ordenación estructurada Contracción de subcampos	54,19% 40,26% Exacta 13,93% Parcial
2. – Equiparación parcial Palabras clave Sin palabras vacías Equiparación parcial Ponderación de los términos de búsqueda (frecuencia inversa)	Palabras incluidas en los encabezamientos Ponderación de los términos de indización (frecuencia inversa)	Encabezamientos de materia Ordenación por relevancia	35,29% 2,32% Exacta 7,04% Parcial 25,93% Algún término

Fig. 5.1 Técnicas de recuperación de información: propuesta de secuenciación

En primer lugar se ejecutaría una búsqueda por equiparación exacta con lectura “secuencial” y truncamiento a la derecha sobre un índice de encabezamientos permutado. De esta manera se recuperarían los encabezamientos que contienen algún subcampo que coincide exactamente y los que “empiezan como” el enunciado. La visualización de los índices alfabéticos de materia puede ayudar a concretar enunciados poco específicos o coincidencias parciales.

Si no hay coincidencia con esta opción se ejecutaría la búsqueda por palabras clave sobre el encabezamiento, eliminando las palabras vacías y con equiparación parcial. De esta forma se recuperarían todos los encabezamientos que contienen los términos del enunciado en cualquier orden dentro de los subcampos sin tener en cuenta los nexos, junto con los que contienen algún término del enunciado. Como el

resultado puede ser muy exhaustivo y los niveles de precisión muy variables, parece recomendable ordenar los encabezamientos por su relevancia en función del número de términos del enunciado que contengan y del número de veces que aparezca el término en la base de datos.

En esta propuesta resulta evidente que, teniendo en cuenta las características de los enunciados y los porcentajes de coincidencia de cada categoría, se ha considerado que la respuesta más adecuada del sistema es la visualización de los índices de materia. Es la opción más directa para proporcionar al usuario la posibilidad de seguir buscando, explorando los índices o consultando la terminología asociada si se aprovecha la estructura sindética de las listas de encabezamientos de materia. Por tanto, cualquier medida que favorezca la presentación de los índices de manera amigable, sintética y clara puede contribuir a mejorar la exploración (Greenberg 1997; Wool 2000; Yee 1999).

A continuación se presentan tres ejemplos de visualización de índices en los que se aplican propuestas de simplificación como la contracción (Cochrane 1986; Allen 1993; Drabenstott y Vizine-Goetz 1994; MacGarry y Svenonius 1991) y la ordenación estructurada (*Guidelines for OPAC displays* 1998; Yee 1999) Además sirven también para complementar el cuadro de la Fig. 5.1 porque serían las la respuesta a tres búsquedas que corresponden a los tres niveles de coincidencia: exacta, parcial y algún término.

El primer ejemplo corresponde a la presentación de una búsqueda por el término *Cine* que coincide exacta y parcialmente con el índice. La presentación sintetiza los 842 encabezamientos recuperados en el catálogo CISNE.

Cine	210 registros	
1 Subdividido por áreas geográficas (Continentes, países, comunidades autónomas, etc.)		Cine de aficionados Cine documental Cine e historia
2 Subdividido por formas de presentación (Diccionarios, enciclopedias, anuarios, etc.)		Cine en color Cine en la literatura Cine Gore
3 Subdividido por periodos cronológicos		Cine negro Cine no profesional Cine y literatura Cine y música Cine y teatro Cine y televisión Cineclubes Cineforum Cines
Cine – Aplicaciones científicas		
Cine – Argumentos y tramas		
Cine – Aspectos antropológicos		
Cine – Aspectos económicos		
Cine – Aspectos educativos		
Cine – Aspectos éticos		
Cine – Aspectos políticos		
Cine – Aspectos sociales		
Cine – Censura		
Cine – Centros de documentación		
Cine – Crítica		
Cine – Derecho		
Cine – Dirección		
Cine – Efectos especiales		
Cine – Efectos sonoros		
Cine – Enseñanza		
Cine – Escenografía y decorados		
Cine – Fotografía		
Cine – Iluminación		
Cine – Lenguaje		
Cine – Producción		
Cine – Semiótica		
Cine – Sonido		
Cine – Técnica		
Cine – Vestuario		
Derechos de autor – Cine		
Pintura – Influencia – Cine		
		Otros términos relacionados
		Dioramas Crítica cinematográfica Directores de cine Documentales Películas cinematográficas
		Se utilizan además los nombres de los distintos géneros cinematográficos: cine negro, comedia cinematográfica, etc.

Fig. 5.2. Ejemplo visualización de índices: coincidencia exacta y parcial

En la parte izquierda aparece en primer lugar la información sobre la coincidencia exacta, indicando el número de registros que tienen ese encabezamiento genérico. También se incluyen aquellos encabezamientos con los que la coincidencia es parcial porque se añaden subdivisiones. Como la búsqueda se realiza sobre un índice permutado, en la parte inferior se sitúan las cadenas en las que el término aparece como subencabezamiento. La forma de visualizar los encabezamientos corresponde a una presentación con contracción y ordenación estructurada⁴.

En la parte derecha se pueden ver aquellos subcampos del índice de materias – en este caso son todos \$a – que empiezan como el enunciado de búsqueda pero tienen más caracteres. Se incluyen también otros términos relacionados con *Cine* en el catálogo de autoridades. Además, la presentación hace transparente al usuario la existencia de relaciones de equivalencia que se activarían automáticamente en caso de seleccionar uno de los encabezamientos relacionados. Por ejemplo, si se selecciona *Cine – Censura* automáticamente se buscará *Películas cinematográficas – Censura* que es el encabezamiento admitido.

El segundo ejemplo corresponde a una búsqueda por el término *Formación*, que coincide sólo parcialmente con el índice y que puede ser tanto la primera palabra de un encabezamiento como de un subencabezamiento. En este caso la presentación sintetiza los 517 encabezamientos recuperados.

⁴ Las características de estas presentaciones pueden verse en el estado de la cuestión, p. 136-139

Formación	
Artes gráficas – Personal – Formación	
Barcos pesqueros – Personal – Formación	
Capital – Formación	Formación continuada
Conceptos – Formación	Formación de anillos
Empresas – Personal – Formación	Formación de palabras
Estrellas – Formación	Formación de personal
Grupos sociales – Formación	Formación de usuarios
Industria del calzado – Personal – Formación	Formación profesional
Industria hotelera – Personal – Formación	Formación profesional y empleo
Metales – Formación	
Profesores – Formación	
Suelos – Formación	
Turismo – Personal – Formación	

Fig. 5.3. Ejemplo visualización de índices: coincidencia parcial

En este ejemplo el término de búsqueda no se utiliza como encabezamiento por lo que en la parte izquierda únicamente aparecen aquellas cadenas en las que se emplea como subdivisión, ordenadas alfabéticamente por el subcampo \$a. En la parte derecha pueden verse de nuevo los subcampos que empiezan con esa palabra (si se selecciona alguno se ampliaría la información). No se incluyen otros términos relacionados porque al emplearse sólo como subencabezamiento no existe información en el catálogo de autoridades.

El último ejemplo es una búsqueda por la expresión *Banco Central Europeo política monetaria*. Puesto que ningún encabezamiento contiene la expresión completa, habría que pasar al segundo nivel de equiparación (Fig. 5.1) y realizar una búsqueda por palabras clave con técnicas de equiparación parcial.

Banco Central Europeo	Consejo Europeo
Política monetaria	Espacio judicial europeo
Banco Europeo de Inversiones	Fondo social europeo
Banco Bilbao Vizcaya	Funcionario europeo
Banco Castellano	Mercado único europeo
Banco Central de Costa Rica	Movimiento europeo
Banco de Andalucía	Parlamentario europeo
Banco de Bilbao	Parlamento Europeo (Estrasburgo)
Banco de España	Sistema europeo de contabilidad
Banco de México	Sistema monetario europeo
Banco de San Carlos (España)	Teatro europeo (Literatura)
Banco óptico	Tribunal Europeo de Derechos Humanos (Estrasburgo)
Banco Santander	Economía monetaria
Comisario europeo	Integración monetaria
Comité económico y social europeo (Bruselas)	Unión monetaria europea

Fig. 5.4. Ejemplo visualización de índices: coincidencia algún término en el índice

En esta presentación los encabezamientos se han ordenado “por relevancia” en función del número de palabras del enunciado que coinciden, su proximidad en la cadena del índice y un “peso” inversamente proporcional a su frecuencia de aparición en el índice. Además, teniendo en cuenta los umbrales de visualización y futilidad que determinan los estudios de Blair (1990) y de Wiberly et al (1995) se ha establecido un valor de corte en 30 encabezamientos.

5.3 Sugerencias para futuras investigaciones.

Para concluir se incluyen algunas sugerencias para futuras investigaciones que han ido surgiendo mientras se realizaba este trabajo. Las propuestas se limitan a señalar brevemente algunas ideas sobre estudios que podrían complementar los resultados de esta investigación.

- ✓ Una línea de trabajo que se perfila claramente es la relacionada con el enriquecimiento del vocabulario de entrada utilizando información de otros campos del registro bibliográfico. Las investigaciones en las que se complementa la información de materias y la de clasificación ya se han realizado en el área anglosajona pero no en catálogos que utilicen listas de encabezamientos en castellano y sistemas de clasificación como la CDU. También sería interesante comprobar la validez de los datos del campo título como para enriquecer la terminología.

- ✓ Otro grupo de estudios que parece necesario estaría relacionado con el comportamiento de búsqueda de los usuarios finales y su opinión sobre los resultados obtenidos empleando diferentes técnicas de recuperación. Se trataría de comprobar, por ejemplo, si se utilizan y cómo los índices o los enlaces hipertextuales de los registros; el efecto de la visualización de diferentes formatos de índices en el comportamiento de búsqueda; o si la utilización de una secuencia de búsqueda como la que se ha propuesto contribuye a mejorar los resultados.

- ✓ Respecto a la existencia de diferencias entre las búsquedas de distintas áreas temáticas que se han verificado, convendría examinar su alcance. Interesaría comprobar si efectivamente los puntos de acceso temáticos de los documentos de Humanidades o Ciencias Sociales son más o menos específicos que los

asignados a los documentos de las áreas “científicas”. También habría que analizar si las evidentes diferencias en el uso de la terminología y en la formulación de expresiones se reproducen en el uso de estrategias de búsqueda (exploración, visualización, selección, etc.) entre usuarios de distintas especialidades. En el caso de que la respuesta sea afirmativa se reforzaría la idea de la necesidad de diseñar interfaces de búsqueda diferentes que se adapten a las necesidades de usuarios diferentes (Hopkins 1993; Ríos García 1991).

6. Bibliografía

Abadal Falgueras, E. "Elementos para la evaluación de consultas de bases de datos web". *El profesional de la información*, 2002, 11(5), p 349-360.

Akeroyd, J. (1990). "Information seeking in online catalogs". *Journal of Documentation*, 1990, 46(1), p.33-52.

The ALA glossary of library and information science. Heartsill Young (ed.). Chicago : American Library Association, 1983.

Allen, B. L. (1991) "Cognitive research in Information Science: implications for desing". En: *Annual Review of Information Science and Technology*. M. E. Willians (ed.) Medford: Information Today, 1991, 26, p. 3-37.

Allen, B. L. (1996) *Information task: toward a user-centered approach to information systems*. San Diego: Academic Press, 1996.

Allen, B. L. (1998) "Visualization and cognitive abilities". En: *Visualizing subject access for Twenty first Century information resources: proceedings of the 1997 Clinic on Library Applications of Data Processing, 2-4 March 1997*. P. A. Cochrane & E. H. Johnson (eds.) Urbana-Champaign, Illinois: Illinois University at Urbana-Champaign, Graduate School of Library and Information Science, 1998, p. 63-79.

Allen, B. L. (2000) "Individual differences and the conundrums of user-centered design: two experiments". *Journal of the American Society for Information Science*, 2000, 51(6), p. 508-520.

Altuna Esteibar, B. (1992). "Comportamientos de uso y estrategias de búsqueda de los usuarios de catálogos automatizados: breve revisión de la investigación". En: *Miscelánea-homenaje a Luis García Ejarque*. Madrid: Federación Española de Sociedades de Archivística, Biblioteconomía y Documentación, 1992 p. 103-111.

Andreu, C.; Escañuela, A.; Torres, C. (1996) "El catàleg de la UAB: una enquesta d'ús". *Item*, 1996, 19, p. 79-93.

Application of MeSH for Medical Catalogers [en línea] <http://www.nlm.nih.gov/tsd/cataloging/catmesh.html> [Consulta: 9 de marzo de 2003]

Association for Library Collections and Technical Services (ALCTS). Cataloging and Classification Section. Subject Access Committee. Subcommittee on Subject Relationships/Reference Structures. *Final report* [en línea]. <<http://www.ala.org/alcts/organization/ccs/sac/rpt97rev.html1997>> [Consulta: 27 de septiembre de 2002]

Atherton, P. (1978) *Books are for use: final report of the Subject Access Project to the Council on Library Resources*. Syracuse, NY: University School of Information Studies, 1978.

Austin, D. ; Digger, J.A. (1977) "PRECIS: the preserved context index system". *Library resources and technical services*, 21(1), 1977, p. 13-30.

- Autoridades de la Biblioteca Nacional* [cd-rom]: Madrid: Ministerio de Educación y Cultura; Chadwycyck-Healy, 1996-
- Baeza-Yates, R.; Ribeiro-Neto, B. (1999) *Modern information retrieval*. New York: ACM Press, 1999.
- Barab, S. A.; Bowdish, B.; Lawless, K. A. (1997) "Hypermedia navigation: profiles of hypermedia users". *Educational Technology Research and Development*, 1997, 45(3), p. 23-41.
- Barry, C. L. (1994) "User-defined relevance criteria: an exploratory study". *Journal of the American Society for Information Science*, 1994, 45(3), p. 149-159.
- Bates, M. J. (1977) "Systems meets user: problems in matching subject search terms". *Information Processing & Management*, 1977, 13, p. 367-375.
- Bates, M. J. (1986) "Subject acces in online catalogs: a desing model." *Journal of the American Society for Information Science*, 1986, 37, p. 357-376.
- Bates, M. J. (1989) "The desing of browsing and berrypicking techniques for the online search interface". *Online Review*, 1989, 13(5), p. 407-424.
- Bates, M. (1990) "Where should the person stop and the information search interface start?". *Information Processing & Management*, 1990, 26(5), p. 575-591.
- Bates, M. (1998) "Indexing and acces for digital libraries and the Internet: human, database and domain factors". *Journal of the American Society for Information Science*, 1998, 49(13), p. 1185-1205.

- Bates, M. J.; Wilde, D. N.; Siegfried, S. (1993) "An analysis of search terminology used by humanities scholars: the Getty Online Searching Project report number 1". *Library Quarterly*, 1993, 63, p. 1-39.
- Beaulieu, M. (1997) "Experiments on interfaces to support query expansion". *Journal of Documentation*, 1997, 53(1), p. 8-19.
- Beaulieu, M.; Borgman, C. (1996) "A new era for OPAC research: introduction to special topic issue on current research in online public access systems". *Journal of the American Society for Information Science*, 1996, 47(7), p. 491-492.
- Beaulieu, M.; Robertson, S.E.; Rasmussen, E.M. (1996) "Evaluating interactive systems in TREC". *Journal of the American Society for Information Science*, 1996, 47, p. 85-94.
- Beheshti, J. (1992) "Browsing through public access catalogs". *Information Technology and Libraries*, 1992, 11(3), p. 220-228.
- Beheshti, J. (1997) "The evolving OPAC". *Cataloging & Classification Quarterly*, 1997, 24(1/2), p. 163-185.
- Beheshti, J., Large, V.; Bialek, M. (1996) "Public Access Catalog Extension (PACE): a browsable graphical interface". *Information Technology and Libraries*, 1996, 15(4), p. 231-240.
- Belkin, N. (1990) "The cognitive viewpoint in Information Science". *Journal of Information Science*, 1990, 16(1), p. 11-16.
- Belkin, N. J.; Croft, W.B. (1987) "Retrieval techniques". *Annual Review of Information Science and Technology*, M. E. Williams (ed.) Medford: Information Today 1987, 22, p. 109-145.

Belkin, N. J.; Oddy, R. N.; Brooks, H. M. (1982). "ASK for information retrieval. Part I: Background and theory". *Journal of Documentation*, 1982, 38(2), p. 61-71.

Belkin, N. J.; Oddy, R. N.; Brooks, H. M. (1982) "ASK for information retrieval. Part II: Results of a desing study". *Journal of Documentation*, 1982, 38(3), p. 145-164.

Belkin, N. J.; Saracevic, T. (1992) "Desing principles for third-generation online public access catalogs: taking account of users and library use". *Annual Review of OCLC Research*, 1992, July 1991-June 1992, p. 43-45.

Belkin, N. J.; et al. (1991) "User interfaces for information systems". *Journal of Information Science*, 1991, 17, p. 327-344.

Belkin, N. J.; et al. (1996) "Using relevance feedback and ranking in interactive searching". En: *Proceedings of the 4th Text REtrieval Conference (TREC-4): 1995 November 1-3, Gaithersburg*. D. K. Hamann (Ed.) Gaithersburg: National Institute of Standards and Technology, 1996, p. 181-209.

Bermello Crespo, L. (2001) "Tendencias de desarrollo de los lenguajes de encabezamientos de materia en las bibliotecas" [en línea]. *Acimed*, 2001, 9(2) <http://bvs.sld.cu/revistas/aci/vol9_2_02/aci03201.htm> [Consulta: 23 de febrero de 2003]

Bes Gracia, M. P.; Romance Carcas, C. (1991) "Recuperación de la información mediante catálogo de acceso público en línea". *Boletín de la ANABAD*, 1991, XLI(3-4), p. 249-256.

Bilal, D. (1998) "Children's search processes in using World Wide Web search engines: an exploratory study". En: *ASIS'98: proceedings of the 61st American Society for Information Science Annual Meeting 35 : October 24-29, 1998, Pittsburg*. R. Larson, K. Petersen, & C. M. Preston (Eds.) Medford, NJ: Information Today, 1998, p. 35-45.

Bilindex 2003: A Spanish-English Subject Heading List. Berkely, Calif. : Floricanto Press, 2003.

Bilindex 2003: A Spanish-English Subject Heading List in the fields of computer, information, communication sciences. Berkely, Calif. : Floricanto Press, 2003.

Blair, D.C. *Language and representation in information retrieval*. Amsterdam: Elsevier, 1990

Blecic, D. D.; et al. (1998) "Using transaction log analysis to improve OPAC retrieval results.". *College & Research Libraries*, 1998, 59(1), p. 39-50.

Blecic, D. D.; et al. (1999) "A longitudinal study of the effects of OPAC screen changes on searching behavior and searches success". *College & Research Libraries*, 1999, 60(6), p. 515-530.

Borgman, C. L. (1986) "The user's mental model of an information retrieval system: an experiment on a prototipe online catalog". *International Journal of Man-Machine Studies*, 1986, 24(1), p. 47-64.

Borgman, C. L. (1986a) "Why are online catalogs hard to use?: lessons learned for information retrieval studies". *Journal of the American Society for Information Science*, 1986, 37(6), p. 387-400.

- Borgman, C. L. (1989) "All users of information retrieval systems are not created equal: an exploration into individual differences". *Information Processing & Management*, 1989, 25(3), p. 237-251.
- Borgman, C. L. (1996) "Why are online catalogs still hard to use?". *Journal of the American Society for Information Science*, 1996, 47(7), p. 493-503.
- Borgman, C. L. (2000) "Why are digital libraries hard to use? In: C. Borgman. *From Gutenberg to the global information infrastructure: access to information in the networked world*. Cambridge: MIT, 2000, p. 117-141
- Borgman, C. L.; Hirsh, S. G.; Hiller, J. (1996) "Rethinking online monitoring methods for information retrieval systems: from search product to search process". *Journal of the American Society for Information Science*, 1996, 47(7), p. 568-583.
- Borgman, C. L.; Hirs, S. G.; Walter, V. A. (1995) "Children's searching behavior on browsing and keyword online catalogs: the Science Library Catalog Project". *Journal of the American Society for Information Science*, 1995, 46(9), p.663-684.
- Borgman, C. L., Moghdam, D.; Corbett, P. K.(1984) *Effective online searching: a basic text*. New York: Marcel Dekker, 1984.
- Borrego Huerta, A. (2001) *Metodología cualitativa de investigación en Biblioteconomía y Documentación: 1993-1999*. Universidad de Salamanca, 2001. Tesis doctoral inédita.
- Bourdon, F. (2001) "Functional requirements and numbering of authority records (FRANAR): to what extent authority control can be supported by technical means?" [en línea]. En: *67th IFLA Council and General Conference: August 16-*

- 25, 2001. <<http://www.ifla.org/IV/ifla67/papers/096-152ae.pdf>> [Consulta: 23 de abril de 2003]
- Brooks, H. M. (1987) "Expert systems and intelligent information retrieval". *Information Processing & Management*, 1987, 23(4), p. 367-382.
- Buckland, M. K. (1995) *Searching multiple digital libraries a desing analysis* [en línea]. Prepared for the Digital Library Project, revised November 2, 1995. <<http://www.sims.berkeley.edu/research/projects/oasis/multisrch.htm#buckland92>> [Consulta: 2 de abril de 2003]
- Buckland, M. (2000) "OASIS [en línea]: studies in information selection: retrieval, filtering & adaptative searching". <<http://www.sims.berkeley.edu/research/oasis>> [Consulta: 26 de septiembre de 2002]
- Buckland, M. K., et al. (1992) "OASIS [en línea]: A Front-end for Prototyping Catalog Enhancements". *Library Hi Tech*, 1992, 10(4), p. 7-22 También disponible en <<http://bliss.berkeley.edu/papers/oasis/oasis.html>> [Consulta: 2 de abril de 2003]
- Buckland, M. K. ;et al. (1999). "Mapping entry vocabularies to unfamiliar metadata vocabularies" [en línea]. *D-Lib Magazine*, 1999, 5(1). <<http://www.dlib.org/dlib/january99/buckland/01buckland.html>> [Consulta: 23 de abril de 2003]
- Byrne, A. and M. Micco. (1988) "Improving OPAC subject access: the ADFA experiment". *College & Research libraries*, 1988, 49, p. 432-441.
- Carlyle, A.(1989) "Matching LCSH and user vocabulary in the library catalog". *Cataloging & Classification Quarterly*, 1989, 10(1/2), p. 37-63.

Caro Castro, C.; Travieso Rodríguez, C. (2003) "Búsquedas por materias en el catálogo virtual: la realidad y el deseo". En: Jornadas Españolas de Documentación Automatizada (8ª. 2003. Barcelona). *Los sistemas de información en las organizaciones: eficacia y transparencia: Fesabid 2003: 8ª Jornadas Españolas de Documentación: Barcelona, 6,7 y 8 de febrero de 2003*. Barcelona: Fesabid : Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 2003, p. 498-499.

Caro Castro, C.; Travieso Rodríguez, C. (2003a) "Encabezamientos de materia en las bibliotecas españolas: prespectiva histórica y situación actual". En: Sociedad Internacional para la Organización del Conocimiento. Capítulo Español. Congreso (6º. 2003. Salamanca). *Tendencias de investigación en organización del conocimiento: actas del VI Congreso ISKO-España, 5-7 de mayo de 2003, Salamanca*. Salamanca: ISKO-España: Departamento de Biblioteconomía y Documentación de la Universidad de Salamanca, 2003, p. 327-334.

Cea D'Ancona, M. A. (1996) *Metodología cuantitativa: estrategias y técnicas de investigación social*. Madrid: Síntesis, 1996.

Chan, L. M. (1992) "Alternatives to subject strings in the Library of Congress Subject Headings system". En: *The future of subdivision in the Library of Congress Subject Headings system: report from the Subject Subdivisión Conference sponsored by the Library of Congress, may 9-12, 1991*. M. O. H. Conway (Ed.) Washington: Library of Congress, Cataloging Distribution Service, 1992.

Chan, L. M. (1995) "Classification, present and future." *Clataloging & Classification Quaterly*,21(2), 1995, p. 5-17.

Chan, L. M. (1995a) *Library of Congress Subject Headings: principles of structure and policies for application*. 3rd ed. Littleton, Colo: Libraries Unlimited, 1995.

Chan, L. M.(2000) "Exploiting LCSH, LCC, and DDC to retrieve networked resources" [en línea]. En: *Confronting the challenges of networked resources and the web: Bicentennial Conference on Bibliographic Control for the New Millennium*. <http://lcweb.loc.gov/catdir/bibcontrol/chan_paper.html> [Consulta: 23 de abril de 2003]

Chan, L. M.; Lin, X.; Zeng, M. (1999) "Un enfoque estructural y multilingüe al acceso por materias en la web" [en línea]. En: *Proceedings from 65th General Council and General Conference: Bangkok, Thailand, August 20-28, 1999*. <<http://www.ifla.org/IV/ifla/papers/o12-117s.htm>> [Consulta: 23 de abril de 2003]

Chan, L. M.; Zeng, M. L. (2002) "Ensuring interoperability among subject vocabularies and knowledge organization schemes: a metodological analysis". *IFLA Journal*, 2002, 28(5/6), p. 323-327.

Chang, S.; Rice, R.E. (1993) "Browsing: a multidimensional framework". En: *Annual Review of Information Science and Technology*. M. E. Williams (ed.) Medford: Information Today, 1993, v. 28, p. 231-276.

Chen, H. (1995). "Machine learning for information retrieval: neural networks, symbolic learning, and genetic algorithms". *Journal of the American Society for Information Science*, 1995, 46(3), p. 194-216.

Chen, H. (1999). "Semantic research for digital libraries"[en línea]. *D-Lib Magazine*, 1999, 5(10). <<http://www.dlib.org/dlib/october99/chen/10chen.html>> [Consulta: 23 de abril de 2003]

Cheng, C.C. (1985) "Microcomputer-based user interface". *Information Technology and Libraries*, 1985, 4(4), p. 346-351.

Cheng, I.; Wilensky, R. (1997) *An experiment in enhancing information access by natural language processing* [en línea]. Computer Science Division at the University of California at Berkeley. <<http://http.cs.berkeley.edu/~wilensky/NLP.html>> [Consulta: 23 de abril de 2003]

Cheung, D. W.; Kao, B.; Lee, J.(1998) "Discovering user access patterns on the World Wide Web.". *Knowledge-Based Systems*, 1998, 10(7), p. 463-470.

Clavel-Merrin, G. (1999) "La necesidad de cooperación en la creación y mantenimiento de archivos temáticos multilingües de autoridades" [en línea].. En: *65th IFLA Council and General Conference: Bangkok, Thailand, August 20-August 28 1999*. <<http://www.ifla.org/IV/ifla65/papers/080-155s.htm>> [Consulta: 23 de abril de 2003]

Clavel-Merrin, G. (2003) "MACS (Multilingual access to subjects) [en línea]: a virtual authority file across languages". En: *Authority control: definition and international experiences: International conference : Florence, Italy, February 10-12 2003*. <<http://www.unifi.it/biblioteche/ac/en/home.htm>> [Consulta: 23 de abril de 2003]

Cleverdon, C. W. (1967) *The Cranfield tests on index language devices*. *ASLIB Proceedings*, 1967, 19(6), p. 173-194.

Coates, E. J. (1960) *Subject catalogues: headings and structure*. London: Library Associationn, 1960.

Cochrane, P. A. (1986) *Improving LCSH for use in online catalogs*, . Littleton, CO: Libraries Unlimited, 1986.

Cochrane, P. A. (2000) "Improving LCSH for use in online catalogs revisited: What progress has been made? What issues still remain?". *Cataloging & Classification Quarterly*, 2000, 29(1/2), p. 73-89.

Cochrane, P. A.; Markey, K. (1983). "Catalog use studies since the introduction of online interactive catalogs: impact on design for subject access". *Library and Information Science Research*, 1983, 5(4), p. 337-363.

Collantes, L. Y. (1995). "Degree of agreement in naming objects and concepts in information retrieval". *Journal of the American Society for Information Science*, 1995, 46(2), p. 116-132.

Colomer, M.; Serra, E. (1992) "L'OPAC a la Biblioteca de Catalunya". En: *OPACs: casos de usuarios de sistemas automatizados de bibliotecas: Jornadas de la SOCADI con la colaboración de DOC6: Asesores en Técnicas de Documentación, Barcelona, 25 de noviembre de 1991*. Barcelona: SOCADI, 1992, p. 55-65.

Computer Science and Telecommunications Board, Commission on Physical Sciences, Mathematics, and Applications, National Research Council. *More than screen deep: toward every-citizen interfaces to the nation's information infrastructure*. Washington D.C.: National Academy Press, 1997.

Conway, M. (ed.). (1992). The future of subdivision in the Library of Congress Subject Headings system: report from the Subject Subdivision Conference sponsored by the Library of Congress, may 9-12, 1991. Washington: Library of Congress, Cataloging Distribution Service, 1992

- Connaway, L. S.; Budd, M.; Kochtanek, T. R. (1995) "An investigation of the use of an online catalog: user characteristics and transaction log analysis". *Library Resources & Technical Services*, 1995, 39(2), p. 142-152.
- Connaway, L. S.; Johnson, D. W.; Searing, S. E. (1997) "Online catalogs from the user's perspective: the use of focus group interviews". *College & Research Libraries*, 1997, 58(5), p. 403-420.
- Connell, T. H. (1995) "Subject searching in online catalogs: metaknowledge used by experienced searchers". *Journal of the American Society for Information Science*, 1995, 46(7), p. 507-518.
- Cool, C. (1993) "Information retrieval as symbolic interaction: examples from humanities scholars". En: *Proceedings of the 56th Annual Meeting of the American Society for Information Science, Columbus, Ohio, 24-28 October 1993*. S. Bonzi (ed.) Medford, New Jersey: Learned Information, 1993, p. 274-277.
- Cooper, M. D. (2001) "Usage patterns of a web based library catalog". *Journal of the American Society for Information Science*, 2001, 52(2), p. 137-148.
- Cove, J. F.; Walsh, B. C. (1988) "Online text retrieval via browsing". *Information Processing & Management*, 1988, 24(1), p. 31-37.
- Coyle, K. (2000). "The virtual union catalog: a comparative study" [en línea]. *D-Lib Magazine*, 2000, 6(3). <<http://www.dlib.org/dlib/march00/coyle/03coyle.html>> [Consulta: 21 de diciembre de 2002]
- Crawford, J. C., Thom; L. C.; Powles, J. A. (1993) "User misconceptions of information retrieval systems". *Journal of Librarianship and Information Science*, 1993, 25(2), p. 85-93.

Crestadoro, A. (1856). *The art of making catalogues of libraries, or, A method to obtain in a short time a most perfect, complete, and satisfactory catalogue of the British Museum Library*, . London: Literary, Artistic Reference Office, 1856.

Croft, W. B.; Thompson, R. H. (1987) "ISR: a new approach to the design of document retrieval systems". *Journal of the American Society for Information Science*, 1987, 38(6), p. 389-404.

Crouch, C. J.; Yang, B. (1992) "Experiments in automatic statistical thesaurus construction". *Communications of the ACM*, 1992, 35(12), p. 77-88.

Cutter, C. A. (1904) *Rules for a dictionary catalog*. 4th ed. Washington: U.S. Government Printing Office, 1904. (Primera ed. 1876)

Dahlberg, I. (1996) "Library catalogs in the Internet: switching for future subject access". En: *Knowledge organization and change: proceedings of the 4th International Isko Conference, 15-18 July 1996, Washington DC, USA*. R. Green (Ed.) Frankfurt: Indeks Verlag, 1996, p. 225-251.

Dalrymple, P. W. (2001) "A quarter century of user-centered study: the impact of Zweizig and Dervin on LIS research". *Library & Information Science Research*, 2001, 23(2), p. 155-165.

Dalrymple, P. W.; Younger, J. J. (1991) "From authority control to information retrieval: framing the expanded domain of subject access". *College and Research Libraries*, 1991, 52(2), p. 139-149.

Dalrymple, P. W.; Zweizig, D. L. (1992) "User's experience of information retrieval systems: an exploration of the relationship between search experience and

- affective measures". *Library & Information Science Research*, 1992, 14(2), p. 167-181.
- Damerau, F. J.A. (1964) "Technique for computer detection and correction of spelling errors". *Communications of the ACM*, 1964, 7(3), p. 171-176.
- Davies, Ch. H.; Rush, J.E. (1979) *Guide to Information Science*. Westport, Conn.: Greenwood Press, 1979.
- De Mey, M (ed.) (1977). *International Workshop on the Cognitive Viewpoint, 1977 March 24-26, Ghent, Belgium*. Ghent, Belgium: University of Ghent, 1977
- Delgado López-Cózar, E. (2000) "Diagnóstico de la investigación en Biblioteconomía y Documentación en España (1976-1996): estado embrionario". *Revista de investigación iberoamericana en ciencia de la información y documentación*, 2000, 1(1), p. 79-93.
- Delgado López-Cózar, E. (2000a) "Los métodos de investigación empleados en la literatura científica producida en Biblioteconomía y Documentación". En: Congreso Universitario de Ciencias de la Documentación (1º. 2000. Madrid). *Teoría, historia y metodología de las Ciencias de la Documentación (1975-2000)*. J. López Yepes (Ed.) Madrid: Universidad Complutense. Departamento de Biblioteconomía y Documentación, 2000, p. 625-641.
- Delsey, T. (2000) "The library catalogue in a networked environment [en línea]". En: *Confronting the challenges of networked resources and the web: Bicentennial Conference on Bibliographic Control for the New Millennium*. <http://lcweb.loc.gov/catdir/bibcontrol/delsey_paper.html> [Consulta: 21 de diciembre de 2002]

Dick, A. (1999) "Epistemological positions and library and information science". *Library Quarterly*, 1999, 67(2), p. 305-323.

Dieberger, Andreas (1997). "Navigation metaphors and social navigation in information spaces" [en línea]. En: *CHI'97 Workshop on Navigation in Information Spaces*.
<http://www.lcc.gatech.edu/~dieberger/CHI97_navigation_workshop.html>
[Consulta: 28 de abril de 2003]

Dillon, A. (2000) "Spatial-semantics: how users derive shape from information space". *Journal of the American Society for Information Science*, 2000, 51(6), p. 521-528.

Dimitroff, A. (1992) "Mental models theory and search outcome in a bibliographic retrieval system". *Library & Information Science Research*, 1992, 14, p. 141-156.

Dimitroff, A.; Wolfram, D. (1995) "Searcher response in a hypertext-based bibliographic Information Retrieval System". *Journal of the American Society for Information Science*, 1995, 46(1), p. 22-29.

Directrices para los registros de autoridad y referencia de materia (1995). Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas. Sección de Clasificación e Indización. Grupo de Trabajo sobre "Directrices para los ficheros de Autoridad de Materia". Madrid: ANABAD; Arco/Libros, 1995.

Dominguez Sanjurjo, R. (1996) *Nuevas formas de organización y servicios en la biblioteca pública*. Gijón: Trea, 1996.

Doszckocs, T. E. (1983) "CITE NLM: natural language searching in an online catalog". *Information Technology and Libraries*, 1983, 2(4), p. 364-380.

Doszkocs, T. E. (1987) "Natural-language searching as a user performance enhancement tool.". *Canadian Journal of Information Science*, 1987, 12(374), p. 139-149.

Drabenstott, K. M.

V.a Markey, K

Drabenstott, K. M. (1994) *Enhancing a new design for subject access to online catalogs*. Ann Arbor, Mich: University of Michigan, School of Information of Library Studies, 1994.

Drabenstott, K. M. (1996) "Authority control in digital libraries" [en línea]. En: *Authority control in the 21st century: an invitational conference : proceedings of the OCLC Symposium (March 31-April 1, 1996)*. <http://www.oclc.org/oclc/man/authconf/index.htm> [Consulta: 24 de abril de 2003]

Drabenstott, K. M. (1996a) "Enhancing a new design for subject access to online catalogs". *Library Hi Tech*, 1996, 14(1), p. 87-109

Drabenstott, K. M.; Simcox, S.; Fenton, E. G. (1999) "End-user understanding of subject headings in library catalogs". *Library Resources & Technical Services*, 1999, 43(3), p. 140-160.

Drabenstott, K. M.; Vazine-Goetz, D. (1990) "Search trees for subject searching in online catalogs". *Library Hi Tech*, 1990, 8(3), p. 7-20.

Drabenstott, K. M.; Vizine-Goetz, D. (1994) *Using subject headings for online retrieval: theory, practice and potential*. San Diego, California: Academic Press, 1994.

Drabenstott, K. M.; Weller, M. S. (1994) "Testing a new desing for subject searching in online catalogs". *Library Hi Tech*, 1994, 12(1), p. 67-76.

Drabenstott, K. M.; Weller, M. S. (1995) *Testing a new desing for subject access to online catalogs*. Ann Arbor, MI: School of Information and Library Studies, 1995.

Drabenstott, K. M., Weller, M. (1996)"A comparative approach to system evaluation: delegating control of retrieval test to an experimental online system". En: *ASIS 1996 Annual Meeting Proceedings: October 19-24, 1996*, <<http://www.asis.org/annual-96/ElectronicProccedings/Dranbenstott/index.html>> [Consulta: 2 de abril de 2003]

Drabenstott, K. M.; Weller, M. S. (1996a) "The exact-display approach for online catalog subject searching". *Information Processing & Management*, 1996, 32(6), p. 719-745.

Drabenstott, K. M.; Weller, M. S. (1996b) "Failure analysis of subject searches in a test of a new desing for subject acces to online catalogs". *Journal of the American Society for Information Science*, 1996, 47(7), p. 519-537.

Drabenstott, K. M.; Weller, M. (1996c) "Handling spelling errors in online catalog searches". *Library Resources & Technical Services*, 1996, 40(2), p. 113-132.

Drabenstott, K. M.; Weller, M. S. (1996d) "Improving personal name searching in online catalogs". *Information Technology and Libraries*, 1996, 15(3), p. 135-155.

Eberts, R. (1994). *User interface design*. Englewood Cliffs : Prentice Hall, 1994.

Efthimiadis, E. N. (1990) *A review of online search aids*. London: British Library, 1990.

Efthimiadis, E. N. (1995) "User choices: a new yardstick for the evaluation of ranking algorithms for interactive query expansion". *Information Processing & Management*, 1995, 31(4), p. 605-620.

Efthimiadis, E. N. (1996) "Query expansion". En: *Annual Review of Information Science and Technology*. M. E. Willians (ed.) Medford: Information Today, 1996. 31, p. 121-187.

Efthimiadis, E. N. (2000) "Interactive query expansion: a user-based evaluation in a relevance feedback environment". *Journal of the American Society for Information Science*, 2000, 51(11), p. 989-1003.

Ellis, D. (1992)"The physical and cognitive paradigms in information retrieval research". *Journal of Documentation*, 1992, 48(1), p. 45-64.

Ellis, D.; Cox, D.; Hall, K. (1993) "A comparison of the information seeking patterns of researches in the physical and social sciences". *Journal of Documentation*, 1993, 49(4), p. 171-212.

Encabezamientos de materia de la Biblioteca Universitaria de Sevilla. Sevilla: Universidad de Sevilla. Secretariado de Publicaciones, 1992 (Suplemento 1994).

Encabezamientos de materia: normativa para su redacción. Madrid: Ministerio de Cultura, 1991.

- Ensor, P. (1992) "User practices in keyword and boolean searching on an online public access catalog". *Information Technology and Libraries*, 1992, 11(3), p. 210-219.
- Espelt, C. (1998) "Improving subject retrieval [en línea]: user-friendly interfaces and effectiveness". *BiD : Biblioteconomia i Documentació*, 1998, 1, <<http://www.ub.es/biblio/bid/01espel1.htm>> [Consulta: 2 de abril de 2003]
- Estivill Rius, A. (1984) "Catàlegs en línia: alguns aspectes de l'accés per matèries". *Bulletí de l'Associació de Bibliotecaris de Catalunya*, 1984, 6, p. 31-62.
- Estivill Rius, A. (1986) "Accés per matèries en els catàlegs en línia d'accés públic". En: *Segones Jornades Catalanes de Documentació, 17-18 de abril, Palau de Congressos Barcelona*. Barcelona: SOCADI, 1986, v. 6, p. 96-105.
- Estivill Rius, A., Gascón, J., Urbano, C. (1995) "Qualitat de la visualització de la informació bibliogràfica en els catàlegs en línia". En: *Jornades Catalanes de Documentació (5^a. 1995. Barcelona). 5es. Jornades Catalanes de Documentació: Biblioteques, centres de documentació i serveis de informació, 25, 26 i 27 d'octubre 1995, Palau de Congressos de Barcelona*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya: Societat Catalana de Documentació i Informació, 1995, p. 481-495.
- Evans, P. (2001) "Z39.50: Part 1- an overview" [en línea]. *Biblio Tech Review*, 2001, <http://www.biblio-tech.com/html/z39_50.html> [Consulta 25 de febrero de 2003]
- Evans, P. (2001) "Z39.50: Part 2- technical details" [en línea]. *Biblio Tech Review*, 2001. <http://www.biblio-tech.com/html/z39_50_part_2.html> [Consulta 25 de febrero de 2003]

- Farradane, J. E. L. (1950) "A scientific theory of classification and indexing and its practical applications". *Journal of Documentation*, 1950, 6(2), p. 83-89.
- Farradane, J. E. L. (1952) "A scientific theory of classification and indexing: further considerations", *Journal of Documentation*, 1952, 8(2), p. 73-92.
- Farradane, J. E. L. (1963) "Relational indexing and classification in the light of recent experimental work in psychology". *Information storage and retrieval*, 1963, 1(1), p. 3-11.
- Fenichel, C. (1981). "Online searching: measures that discriminate among users with different types of experience". *Journal of the American Society for Information Science*, 1981, 23, p. 23-32.
- Ferl, T. E.; Millsap, L. (1996) "The Knuckle-Cracker's dilemma: a transaction log study of OPAC subject searching". *Information Technology and Libraries*, 1996, 15(2), p. 81-98.
- Fernández Molina, J. C.; Moya Anegón, F. de (1998) *Los catálogos de acceso público en línea: el futuro de la recuperación de información*. Málaga: Asociación Andaluza de Bibliotecarios, 1998.
- Fernández Molina, J. C.; Moya Anegón, F. de (2002). "Perspectivas epistemológicas "humanas" en la documentación". *Revista española de documentación científica*, 2002, 25(3), p. 241-253.
- Fernández Molina, J. C.; Peis, E. (1995) "Análisis de los métodos utilizados en la mejora de la calidad del acceso por materias en los OPAC's". En: Jornades Catalanes de Documentació (5^a. 1995. Barcelona) 5es. *Jornades catalanes de Documentación: Biblioteques, centres de documentació i serveis de informació*, 25, 26 i 27 d'octubre 1995, Palau de Congressos de Barcelona. Barcelona:

- Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya: Societat Catalana de Documentació i Informació, 1995, p. 471-479.
- Fernández Molina, J. C.; Peis, E. (1997) "Búsqueda dinámica y exploratoria: el browsing en los catálogos en línea". *Investigación bibliotecológica*, 1997, 22, p. 28-38.
- Fidel, R. (1991) "Searches selection of search keys. I The selection routine". *Journal of the American Society for Information Science*, 1991, 42(7), p. 490-500.
- Fidel, R. (1991) "Searches selection of search keys. II Controlled vocabulary or free-text searching". *Journal of the American Society for Information Science*, 1991, 42(7), p. 501-514.
- Fidel, R. (1991) "Searches selection of search keys. III Searching styles". *Journal of the American Society for Information Science*, 1991, 42(7), p. 515-527.
- Fidel, R. (1993) "Qualitative methods in information retrieval research". *Library & Information Science Research*, 1993, 15(3), p. 219-247.
- Fidel, R. (2000) "The user centered approach: how we got here". En: *Saving the time of the library user through subject access innovation: papers in honor of Pauline Atherton Cochrane*. W. J. Wheeler (Ed.) Champaign, IL: Graduate School of Library and Information Science, 2000, p.79-99.
- Fidel, R.; Soergel, D. (1983) "Factors affecting online bibliographic retrieval: a conceptual framework for research". *Journal of the American Society for Information Science*, 1983, 34, p. 163-180.

Fidel, R.; et al. (1999) "A visit to the information mall: Web searching behavior of high school students". *Journal of the American Society for Information Science*, 1999, 50(1), p. 24-37.

Ford, N. (2000) "Cognitive styles and virtual environments". *Journal of the American Society for Information Science*, 2000, 51(6), p. 543-557.

Formato IBERMARC para registros bibliográficos (2002). 6ª ed. Madrid : Ministerio de Cultura, Biblioteca Nacional, 2002

Formato IBERMARC para registros de autoridad (1999). Madrid : Biblioteca Nacional, 1999

Formato MARC21 conciso para información de la comunidad [en línea]. Biblioteca del Congreso, Oficina de desarrollo de redes y normas MARC. <<http://www.loc.gov/marc/community/eccispa.html>> [Consulta: 25 de febrero de 2003]

Foskett, A. C. (1997) *The subject approach to information*. 5th. ed. London: Library Association, 1997.

Fox, E. A.; et al. (1993). "Development of a Modern OPAC: from REVTOC to MARIAN". En: *SIGIR'93: Proceedings of the Association for Computing Machinery Special Interest Group in Information Retrieval (ACM/SIGIR) 16th Annual International Conference on Research and Development in Information Retrieval: 1993, June 27-July 1, Pittsburg, PA*. R. Korfhage, E. Rasmussen, & P. Willett (Eds.) New York: ACM, 1993, p. 248-259.

Frakes, W.B. (1992) "Stemming algorithms". En : Frakes, W.B.; Baeza-Yates, R. (eds.). *Information retrieval: data structures and algorithms*. Englewood Cliffs: Prentice Hall, 1992, p. 131-160.

- Frakes, W.B.; Baeza-Yates, R. (eds.) (1992). *Information retrieval: data structures and algorithms*. Englewood Cliffs: Prentice Hall, 1992
- Franz, L.; et al. (1994) "End-user understanding of subdivided subject headings". *Library Resources & Technical Services*, 1994, 38(3), p. 213-226.
- Freeman, R. R.; Atherton, P. (1968) "File organization and search strategy using the Universal Decimal Classification in mechanized reference retrieval systems". En: *Mechanized information storage, retrieval and dissemination: proceedings of the FID/IFIP Joint Conference, Rome, June 14-17, 1967*. K. Samuelson. Amsterdam: North-Holland, 1968, p. 122-152.
- Frías, J. A. (1994) "El control de autoridades y el acceso a la información". En: *Catalogación de documentos teoría y práctica*. M. Pinto Molina (ed.) Madrid: Síntesis, 1994, p. 435-477.
- Frías, J. A. (1997) "La utilización del catálogo por los usuarios de la biblioteca: luces y sombras". *Revista general de información y documentación*, 1997, 7(1), p. 231-252.
- Frías, J. A.; Martín Rodríguez, F. (1999) "El análisis transaccional como técnica de recogida de datos para el estudio del comportamiento de los usuarios del catálogo en línea". En: *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de información: Actas del IV Congreso ISKO-España ECONSID'99*. López Huertas, M.J., Fernández Molina, J. C. (eds.). Granada, 1999, p. 427-434.
- Frías, J. A.; Ríos Hilario, A. B. (2002) "Visibility and invisibility of the kinship relationships in bibliographic families of the library catalogue". En: *Challenges in Knowledge representation and organization for the 21st Century: integration*

knowledge across boundaries: proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain. M. J. López-Huertas (Ed.) Würzburg: Ergon, 2002, p.264-270.

Fugmann, R. (1993) *Subject analysis and indexing: theoretical foundation and practical advice.* Frankfurt/Main: Indeks Verlag, 1993.

Functional requirements for bibliographic records [en línea]: *final report.* (1998) Approved by the Standing Committee of the IFLA Section on Cataloging. München: Saur, 1998. También disponible en: <<http://www.ifla.org/VII/s13/wgfrbr/wgfrbr.htm>> [Consulta: 25 de febrero de 2003]

García Camarero, E.; García Melero, L. A. (2001) *La biblioteca digital.* Madrid: Arco/Libros, 2001.

García Marco, F. J. (1994). "Hypertexto y lenguajes documentales: retos y sinergías". En: Jornadas Españolas de Documentación Automatizada (4ª. 1994. Gijón). *Los profesionales ante el reto del siglo XXI: integración y calidad: actas, IV Jornadas Españolas de Documentación Automatizada DOCUMAT 94 (Gijón, 6, 7 y 8 de octubre de 1994).* Oviedo, Universidad, 1994, p. 417-426.

García Marco, F. J. (1995) "Interfaces amigables para la recuperación de la información bibliográfica". *Scire*, 1995, 1(1), p. 127-148.

Gardner, S. A. (1992) "Spelling errors in online databases: what the technical communicator should know". *Technical communication*, 1992, 39(1), p. 50-55.

Gaynor, E. (1996) *From MARC to MARKUP* [en línea]: *SGML and online library systems.* <<http://xml.coverpages.org/gaynorMARC96.html>> [Consulta: 20 de marzo de 2003]

- Gil Leiva, I. (2001) "Consistencia en la asignación de materias en las Bibliotecas Públicas del Estado". *Boletín de la Asociación Andaluza de Bibliotecarios*, 2001, 63, p. 69-86.
- Gonçalves, M. A.; France, R. K.; Fox, E. A. (2001) "MARIAN [en línea]: a flexible Interoperability for Federated Digital Libraries". En: *Research and Advanced Technology for Digital Libraries: 5th European Conference, ECDL-01 (Darmstadt, Germany: 4-9 sept. 2001)* <<http://www.dlib.vt.edu/products/Marian.html>> [Consulta 27 de marzo de 2003]
- Gödert, W. (1991) "Facet classification in online retrieval." *International Classification*, 1991, 12(2), p. 88-109.
- Gorman, M. (1992). "After AACR2R: the future of the Anglo-American Cataloging Rules". En: *Origins, content and future of the AACR2 revised ed.* R. P. Smiraglia. Chicago, American Library Association, 1992, p. 89-94.
- Gorman, M. (2000) "From card catalogues to WebPACS [en línea]: celebrating cataloguing in the 20th century". En: *Confronting the challenges of networked resources and the web: Bicentennial Conference on Bibliographic Control for the New Millennium.* <http://lcweb.loc.gov/catdir/bibcontrol/gorman_paper.html> [Consulta: 21 de diciembre de 2002]
- Gonzalo, J.; Peñas, A.; Verdejo, F. (2002) "La indexación con técnicas lingüísticas en el modelo clásico de recuperación de información". En: Jornadas de Tratamiento y Recuperación de Información (1ª. 2002. Valencia). *Actas de las I Jornadas de Tratamiento y Recuperación de Información (JOTRI): Facultad de Informática Valencia, 4 y 5 de julio de 2002.* E. Sanchis, L. Moreno, & I. Gil (Eds.) Valencia: Universidad Politécnica de Valencia, 2002, p. 97-106.

Greenberg, J. (1997) "Reference structures: stagnation, progress, and future challenges". *Information Technology and Libraries*, 1997, 16, p. 108-119.

Greene, D. J. (1996) "INNOPAC Millenium: preparing libraries for the 21st century: a pesrpective and comminment". *Library Hi Tech*, 1996, 14(4), p. 45-56.

Guenther, R. S. (1996) "Bringing the Library of Congress Classification into the computer age: converting LCC to machine readable form". En: *Knowledge organization and change: proceedings of the 4th International Isko Conference, 15-18 July 1996, Washington DC, USA*. R. Green (Ed.) Frankfurt: Indeks Verlag, 1996, p. 26-32.

Guerrero, V.; Moya-Anegón, F. de; Herrero-Solana, V. (2002) "Document organization using Kohonen's algorithm". *Information Processing & Management*, 2002, 38(1), p. 79-89.

Guerrero, V.; et. al. (2002) "Clasificación de términos mediante el algoritmo de Kohonen". En: *Jornadas de Tratamiento y Recuperación de Información (1ª. 2002. Valencia). Actas de las I Jornadas de Tratamiento y Recuperación de Información (JOTRI): Facultad de Informática Valencia, 4 y 5 de julio de 2002*. E. Sanchis, L. Moreno, & I. Gil (Eds.) Valencia: Universidad Politécnica de Valencia, 2002, p. 63-70.

Guidelines for authority records and references (2001) 2nd ed. revised by the IFLA Working Group on GARE revisión. München: KG Saur, 2001

Guidelines for OPAC displays (1998). Prepared for the IFLA task Force on Guidelines for OPAC Displays by Martha Yee. Draft version. November 24, 1998.

- Hancock-Beaulieu, M. (1987). "Subject searching behavior at the library catalog and at the selves: implications for online interactive catalogues". *Journal of Documentation*, 1987, 43(4), p. 303-321.
- Hancock-Beaulieu, M. (1990) "Evaluating the impact of an online library catalogue on subject searching behaviour at the catalogue and at the shelves". *Journal of Documentation*, 1990, 46(4), p. 318-338.
- Hancock-Beaulieu, M. (1993) "A comparative transaction log analysis of browsing and search formulation in online catalogues". *Program*, 1993, 27(3), p. 269-280.
- Hancock-Beaulieu, M.; Fieldhouse, M.; Do, T. (1995) "An evaluation of interactive query expansion in an online library catalogue with a graphical user interface". *Journal of Documentation*, 1995, 51(3), p. 225-243.
- Hancock-Beaulieu, M., Robertson, S. E.; Neilson, C. (1991). "Evaluation of online catalogues: eliciting information from the user". *Information Processing & Management*, 1991, 27, p. 523-532.
- Hancock-Beaulieu, M.; et al. (1995) "An evaluation of interactive query expansion in an online library catalogue with a graphical user interface". *Journal of Documentation*, 1995, 51, p. 225-243.
- Hanson, E. R.; Daily, J. E. (1970) "Catalogs and cataloging". En: *Encyclopedia of Library and Information Science*. A. Kent & H. Lancour (eds.) New York: Marcel Dekker, 1970, v. 4, p. 242-305.
- Harman, D. (1992) "Ranking algorithms". En : Frakes, W.B.; Baeza-Yates, R. (eds.). *Information retrieval: data structures and algorithms*. Englewood Cliffs: Prentice Hall, 1992, p. 363-392

Harman, D. (1992a) "Relevance feedback and other query modification techniques". En : Frakes, W.B.; Baeza-Yates, R. (eds.). *Information retrieval: data structures and algorithms*. Englewood Cliffs: Prentice Hall, 1992, p. 241-263

Harman, D. (1994) "Automatic indexing". En: *Challenges in indexing electronic text and images*. R. Fidel et al. (eds.) Medford:: Learned Information, 1994, p. 247-264.

Harman, D. (1995) "Overview of the Third Text Retrieval Conference (TREC-3)". En: *Overview of the Third Text Retrieval Conference (TREC-3)*. D. Harman (ed.). Gaithersburg, MD, NIST, 1995; p. 1-19.

Harrison, A. D., Roos, F. A., Thomas, R. E. (1995) "(Semi) automatic capturing of bibliographic information from journal contents pages for inclusion in online library catalogues: the RIDDLE Project". *The Electronic Library*, 1995, 13(1), p. 15-20.

Hart, F. E. ; R. Reitsma. (1990) "Subject searching and tables of contents in single-work titles". *Technical Services Quarterly*, 1990, 7(1), p. 33-51.

Harter, S. P. (1992) "Psychological relevance and Information Science". *Journal of the American Society for Information Science*, 1992, 43(9), p. 602-615.

Harter, S. P.; Hert, C. A. (1997) "Evaluation of information retrieval systems: approaches, issues, and methods". *Annual Review of Information Science and Technology*, M. E. Williams (ed.) Medford: Information Today, 1997, 32, p. 3-93.

Headings for tomorrow: public access display of subject headings (1992). Prepared by the Subcommittee on the Display of Subject Headings in Subject Indexes in

- Online Catalogs, Subject Analysis Committee, Cataloging and Classification Section, Association for Library Collections and Technical Services. Chicago; London: American Library Association, 1992.
- Heiner-Freiling, M. (2000) "Survey on subject heading languages used in national libraries and bibliographies". *Cataloging & Classification Quarterly*, 2000, 29(1/2), p. 189-198.
- Helander, M. G.; Landauer, T. K. ; Prabhu, P. V. (eds.) (1997) *Handbook of human-computer interaction*. Amsterdam: Elsevier, 1997.
- Hemmje, M. (1995) "Lyberworld - a 3D graphical user interface for fulltext retrieval" [en línea] . En: *Pro-ceedings of CHI'95*. <<ftp://ftp.darmstadt.gmd.de/pub/VISIT/papers/hemmje/CHI95.ps.gz>> [Consulta: 28 de abril de 2003]
- Hendley, R.; et al. (1995). *Narcissus: visualising information* [en línea]. <<ftp://ftp.cs.bham.ac.uk/pub/authors/R.J.Hendley/ieeeviz.ps.Z>> [Consulta: 28 de abril de 2003]
- Herner, S. (1970) "Browsing". En: *Encyclopedia of Library and Information Science*. A. Kent and H. Lancour. New York: Marcel Dekker, 1970, 3, p. 408-415
- Hert, C. A. (1996) "User goals on an online public acces catalog". *Journal of the American Society for Information Science*, 1996, 47(7), p. 504-518.
- Hert, C. A. (1997) *Understanding information retrieval interactions*. Greenwich: Ablex, 1997.

Hildreth, C. R. (1984) "Pursuing the ideal: generation of online catalogs". En: *Online catalogs, online reference: converging trends*. B. Aveney, B. Butler (eds.) Chicago: American Library Association, 1984.

Hildreth, C. R. (1985) "Online public access catalogs". En: *Annual Review of Information Science and Technology*. M. E. Willians (ed.). White Plains, NY: Knowledge Industry Publications, 1985, v. 20, p. 233-285.

Hildreth, C. R. (1987) "Beyond boolean: designing the next generation of online catalogs". *Library Trends*, 1987, 35, p. 647-667.

Hildreth, C. R. (1989) *Intelligent interfaces and retrieval methods for subject searching in bibliographic retrieval systems*. Washington, D.C., Cataloging Distribution Service, Library of Congress, 1989

Hildreth, C. R. (1989a) "OPAC research: laying the groundwork for future OPAC desing". En: *The online catalog: development and directions*. C. R. Hildreth (ed.) London: The Library Association, 1989, p. 1-24.

Hildreth, C. R. (1991) "Ends users and structured searching of online catalogues: recent research findings". En: *Advances in information organization: tools for Knowledge Organization and the human interface: proceedings of the 1st International-ISKO Conference held in Darmstadt, 14-17 August 1990*. e. b. R. Fugmann (Ed.) Frankfurt/Main: Indeks Verlag, 1991, v. 2, p. 11-15

Hildreth, C. R. (1995) "The GUI OPAC [en línea]: approach with caution". *The Public Access Computer Systems Review*, 1995, 6(5).
<<http://info.lib.uh.edu/pr/v6/n5/hild6n5.html>> [Consulta: 3 de diciembre de 2002]

- Hildreth, C. R. (1995a) *Online catalog desing models [en línea]: are we moving in the right direction?: a report submitted to the Council on Library Resources August, 1995* . <<http://phoenix.liunet.edu/~hildreth/clr-opac.html>> [Consulta: 25 de febrero de 2003]
- Hildreth, C. R. (1997) "The use and understanding of keyword searching in a university online catalog". *Information Technology and Libraries*, 1997, 16(2), p. 52-62.
- Hildreth, C. R. (2001) "Accounting for users' inflated assessments of on-line catalogue search performance and usefulness [en línea]: an experimental study". *Information research*, 2001, 6(2). <<http://informationr.net/ir/6-2/infres62.html>> [Consulta: 25 de febrero de 2003]
- Hirsh, S. G. (1997) "How do children find information on different types of task?: Children's use of the science library catalog". *Library Trends*, 1997, 45, p. 725-745.
- Hjerpe, R. (1989) "HYPERCAT at LIBLAB in Sweden: a progress report". En: *The online catalogue: development and directions*. C. R. Hildreth. London, The Library Association, 1989 p. 177-209.
- Hjorland, B. (2000) "Library and Information Science: practice, theory, and philosophical basis". *Information Processing & Management*, 2000, 36(3), p. 501-531.
- Hodge, G. (2000) *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington, DC: The Digital Library Federation, Council on Library and Information Resources, 2000.

Holm, L.A. (1999) "ONE Project-results and experiences". *International Cataloging and Bibliographic Control*, 1999, 28(1), p. 29-33.

Honkela, T.; et al. (1996). "Self-organizing maps of document collections" [en línea]. *Alma*, 1(2), 1996. <<http://www.diemme.it/~luigi/websom.html>> [Consulta: 28 de abril de 2003]

Honrado, A.; et al. (2002) "A Word Stemming Algorithm for the Spanish Language" [en línea] En: *Seventh International Symposium on String Processing Information Retrieval (SPIRE'00) : September 27 - 29, 2000 A Coruña, Spain*. <<http://www.computer.org/proceedings/spire/0746/07460139abs.htm>> [Consulta: 21 de abril de 2003]

Hopkins, J. (1993) "The one-stop information store: the catalog beyond Cutter". *Cataloging & Classification Quarterly*, 1993, 16(1), p. 127-138.

Hsieh-Yee, I. (1993) "Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers". *Journal of the American Society for Information Science*, 1993, 44(3), p. 161-174.

Hunter, R. N. (1991). "Successes and failures of patrons searching the online catalog at a large academic library: a transaction log analysis". *RQ*, 1991, 30(3), p. 395-402.

Husain, S.; O'Brien, A. (1992) "Recent trend in subject access to OPACs: an evaluation". *International Classification*, 1992, 19(3), p. 140-145.

Information Infrastructure Technology and Applications Workshop [en línea]. <<http://www.iita.ucar.edu/ws/datawksshop/Agenda.html>> [Consulta: 2 de abril de 2003]

Ingwersen, P. (1992) *Information retrieval interaction*. London: Taylor Graham, 1992.

Ingwersen, P. (1996) "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory". *Journal of Documentation*, 1996, 52(1), p. 3-50.

IODyne overview [en línea]. <<http://www.canis.uiuc.edu/news/iodyne.html>>
[Consulta: 25 de abril de 2003]

Irgens, B. (2000) "Problems using a Web-OPAC: a pilot study". *Proceedings of the ASIS Annual Meeting*, 2000, 37, p. 93-108.

ISO/IEC 13250 *Topic Maps* [en línea].
<<http://www.y12.doe.gov/sgml/sc34/document/iso13250-2nd-ed-v2.pdf>>
[Consulta: 23 de abril de 2003]

Iyer, H. (1995) *Classificatory structures: concepts, relations and representation*. Frankfurt/Main: Indeks, 1995.

Iyer, H. (1995a) "Online searching: use of classificatory structures". En: *Advances in information organization: tools for Knowledge Organization and the human interface: proceedings of the 1st International-ISKO Conference held in Darmstadt, 14-17 August 1990*. e. b. R. Fugmann, ed. Frankfurt/Main: Indeks Verlag, 1990, v. I, p. 159-167.

Jacso, P. (1998) "Post-search aids for relevance ranking of results by the user in the Experimental Search System of the Library of Congress". En: *Annual National Online Meeting, May 12-14 1998, New York, USA.* , 1998, p. 212-223.

Jamieson, S. H.; Oddy, R. N. (1979) *Implementation and evaluation of interactive retrieval through an intelligent terminal: a project proposal to the British Library Research and Development Department*. Birmingham: University of Aston in Birmingham, Computer Centre, 1979.

Jansen, B. J.; Spink, A.; Saracevic, T. (2000) "Real life, real users, and real needs: a study and analysis of user queries on the web". *Information Processing & Management*, 2000, 36(2), p. 207-227.

Järvelin, K.; Vakkari, P. (1990) "Content analysis of research articles in Library and Information Science". *Library and Information Science Research*, 1990, 12, p. 395-421.

Jimenez Pelayo, J.; García Blanco, R. (2002) *El catálogo de autoridades: creación y gestión en unidades documentales*. Gijón: Trea, 2002.

Jiménez Rodríguez, J. E. (1998) "Consideraciones sobre la indización en las bibliotecas universitarias españolas". En: Jornadas Españolas de Documentación (6ª. 1998. Valencia). *Los sistemas de información al servicio de la sociedad: actas de las Jornadas, Valencia 1998*. Valencia: FESABID, 1998, p. 461-468.

Jing, Y., Croft, W. B. (1994) "The association thesaurus for information retrieval". En: *RIAO'94: Conference proceedings with presentation of prototypes and operational systems, intelligent multimedia information retrieval system and management: 1994 October 11-13*. New York; Paris: CID, 1994, p. 146-160.

Johnson, A. ; Fotouhi, F. (1998) "Adaptative clustering of hypermedia documents". En: *Encyclopedia of Library and Information Science*. A. Kent and C. M. Hall. New York: Marcel Dekker, 1998. v. 63(26), p. 1-18.

- Joint, N. (2001) "Design interfaces for distributed electronic collections: the lessons of traditional librarianship". *Libri*, 2001, 51, p. 148-156.
- Jokic, M. (1997) "Analysis of users' searches of CD-ROM databases in the National and University Library in Zagreb". *Information Processing & Management*, 1997, 33(6), p. 785-802.
- Jones, R. (1986) "Improving Okapi: transaction log analysis of failed and searches in an online catalogue". *Vine*, 1986, 62, p. 3-13.
- Jones, S.; et al. (1997) "Transaction logging". *Journal of Documentation*, 1997, 53(1), p. 35-50.
- Jones, S.; Cunningham, S. J.; McNab, R. (1998) "Usage analysis of a digital library". En: *Proceedings of the Third ACM Conference on Digital Libraries, Pittsburg, PA.*, 1998, p. 291-295.
- Joy, F. (1995) "Extending the third generation OPAC [en línea]". *ITs News, Journal of the Library Association Information Technology Group*, 1995, 32 p. 27-32 (También publicado como: Fretwell-Downing Informatics. *Projects: papers and discussion documents* <<http://www.fdggroup.co.uk/FDI/x3gopac.htm>> [Consulta: 2 de abril de 2003])
- Kaiser, J. O. (1911) *Systematic indexing* . London: Gibson, 1911.
- Knapp, P. B. (1944) "The subject catalog in the college library". *Library Quarterly*, 1944, 14, p. 214-228.
- Kaske, N. K. (1988). "The variability and intensity overtime of subject searching in an online public access catalog". *Information Technology and Libraries*, 1988, 7(3), p. 273-87.

Kaski, S.; et al. (1996). "Creating an order in digital libraries with self-organizing maps" [en línea]. En: *Proceedings of World Congress on Neural Networks, WCNN '96*. Mahwah, NJ : INNS Press, 814-817. <<http://websom.hut.fi/websom/doc/wcnn96o.ps.gz>> [Consulta: 28 de abril de 2003]

Kiestra, M. D.; Stokmans, M. J. W.; Kamphuis, J. (1994) "End-users searching the online catalog: the influence of domain and system knowledge on search patterns". *The Electronic Library*, 1994, 12, p. 335-343.

Khoo, C. S.G.; Poo, D. C. C. (1994). "An expert system approach to online catalog subject searching". *Information Processing & Management*, 1994, 30(2), p. 223-238.

Kim, Y.; Norgard, B. (1998) *Adding natural language processing techniques to the entry vocabulary module building process* [en línea]. <<http://www.sims.berkeley.edu/research/metadata/papers/nlptech.html>> [Consulta: 2 de abril de 2003]

Kohonen, T. (1997) *Self-organizing maps*. 2nd ed. Berlín : Springer, 1997.

Kohonen, T.; et al. (1999) "Self organization of a massive text document collection". En: *Kohonen Maps*. E. Oja y S. Kaski (Eds.) Amsterdam: Elsevier, 1999, p. 171-182.

Kuhlthau, C.; et al. (1992) "Exploration into stages in the information search process in online information retrieval: communication between users and intermediaries". En: *Proceedings of the 55th Annual Meeting of the American Society for Information Science, Pittsburgh, 26-29 Oct 92*. D. Shaw (Ed.) Medford, New Jersey: Learned Information, 1992, v. 29, p.67-71.

- Kuhlthau, C. C. (1991) "Inside the search process: information seeking from the user's perspective". *Journal of the American Society for Information Science*, 1991, 42, p. 361-371.
- Kuhlthau, C. C. (1993) " A principle of uncertainty for information seeking". *Journal of Documentation*, 1993, 49(4), p. 339-355.
- Kuhlthau, C. C. (1998) "An uncertainty principle for information seeking: a qualitative approach". En: *Encyclopedia of library and information science*. A. Kent (ed.) New York: Marcel Dekker, 1998, v. 61(24), p. 355-368.
- Lagus, K. ; et al. (1996). "Self-organizing maps of document collections: a new approach to interactive exploration" [en línea]. En: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Simoundis, E.; Han, J.; Fayyad, U. (eds.) Menlo Park : AAAI Press, 1996. <<http://websom.hut.fi/websom/doc/kdd96.ps.gz>> [Consulta: 28 de abril de 2003]
- Lancaster, F.W. (1996) *Indización y resúmenes: teoría y práctica*. Buenos Aires: EB publicaciones, 1996.
- Lancaster, F. W. (2002) *El control del vocabulario en la recuperación de información*. 2ª ed. rev. Valencia: Universitat, 2002
- Lancaster, F. W.; et al. (1991) "Identifying barriers to effective subject access in library catalogs". *Library Resources & Technical Services*, 1991, 35, p. 377-391.
- Lancaster, F. W.; et al. (1994) "Searching databases on CD-ROM: comparison of the results of end-user searching with results from two modes of searching by skilled intermediaries". *RQ*, 1994, 33(3), p. 370-386.

- Large, A.; Beheshti, J. (1997) "OPAC's: a research review". *Library and Information Science Research*, 1997, 19(2), p. 111-133.
- Larson, R. (1986) *Workload characteristics and computer systems utilization in online library catalogs*, . Ph. doctoral dissertation, University of California, Berkeley, 1986.
- Larson, R. (1991) "Between Scilla and Charybdis: subject searching in the online catalog". *Advances in Librarianship*, 1991, 15, p. 175-236.
- Larson, R. (1991a)"Classification clustering, probabilistic information retrieval and the online catalog". *Library Quarterly*, 1991, 61(2), p. 133-173.
- Larson, R. (1991b). "The decline of subject searching: long-term trends and patterns of index use in an online catalog". *Journal of the American Society for Information Science*, 1991, 42(3), p. 197-215
- Larson, R. (1992). "Evaluation of avanced retrieval techniques in an experimental online catalog". *Journal of the American Society for Information Science*, 1992, 43(1), p. 34-53.
- Larson, R.; et al. (1996) "Chesire II: designin a next-generation online catalog". *Journal of the American Society for Information Science*, 1996, 47(7), p. 555-567.
- Lawrence, G. S. (1985) "System features for subject access in the online catalog". *Library Resources & Technical Services*, 1985, 29, p. 16-33.
- Lazonder, A. W.; Biemans, H. J. A.; Wopereis, I. G. (2000) "Diferences between novice and experienced users in searching information on the World Wide Web".

Journal of the American Society for Information Science, 2000, 51(6), p. 576-581.

Lester, M. A. (1989). *Coincidence of user vocabulary and Library of Congress Subject Headings: experiments to improve subject access in academic library online catalogs*. Ph. D. dissertation. University of Illinois at Urbana-Champaign, 1989.

Library of Congress subject headings (2002) 25th ed. Washington : Library of Congress, Cataloging Distribution Service, 2002

Lipetz, B. A.; Paulson, P. J. (1987). "A study of the impact of introducing an online subject catalog at the New York State Library". *Library Trends*, 1987, 35(4), p. 597-617.

Lista de encabezamientos de materia para bibliotecas públicas. 2ª ed. rev. Madrid: Ministerio de Cultura, 1995.

Logan, E. (1990) "Cognitive styles and online behavior of novices searches". *information Processing & Management*, 1990, 26, p. 503-510.

Long, C. E. (2000) "Improving subject searching in Web-based OPAC's: evaluation of the problem and guidelines for desing". *Journal of Internet cataloging*, 2000, 2(3/4), p. 159-186.

López de Sosoaga Torija, A (1998) "Opac-Web-Z39.50 [en línea]: ¿Redundantes o complementarios? La realidad es multilingüe". En: *Jornadas Españolas de Documentación (6ª. 1998. Valencia). Los sistemas de información al servicio de la sociedad: actas de las Jornadas*. Valencia: FESABID, 1998. <<http://lcweb.loc.gov/z3950/agency/papers/documat3.doc>> [Consulta: 21 de diciembre de 2002]

López de Sosoaga Torija, A. (1999) "Z39.50 en el siglo XXI [en línea]: ¿estándar real o virtual?". En: *7es. Jornades catalanes de Documentación: EXPODOC'99: Las bibliotecas y los centros de documentación en el siglo xxi: pieza clave de la sociedad de la información, Barcelona 4, 5 y 6 de noviembre de 1999*. Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya: Societat Catalana de Documentació i Informació, 1999. <<http://www.cobdc.org/09jornades/7es/76.pdf>> [Consulta: 21 de diciembre de 2002]

López-Huertas, M. J. (1997) "Thesaurus structure desing: a conceptual approach for improved interaction". *Journal of Documentation*, 1997, 53(2), p. 139-177.

López-Huertas, M.J. (1998) "El diseño de tesauros: consideraciones teóricas y metodológicas". En: *II Encuentros sobre Organización del conocimiento en sistemas de Información y Documentación*. Zaragoza: Universidad; Isko, 1998, p. 67-73.

López-Huertas Pérez, M.J. (1998a) "La representación del usuario en la recuperación de la información". En: *Jornadas Españolas de Documentación (6ª. 1998. Valencia)*. *Los sistemas de información al servicio de la sociedad: actas de las Jornadas*. Madrid: FESABID, 1998, p. 521-527.

López-Huertas, M. J. (1999) "Potencialidad evolutiva del tesoro: hacia una base de conocimiento experto". En: *Sociedad Internacional para la Organización del Conocimiento. Capítulo Español. Congreso (4º. 1999. Granada) La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información: actas del IV Congreso ISKO-España EOCONSID '99, 22-24 de abril de 1999, Granada*. López Huertas, M.J., Fernández Molina, J. C. (eds.). Granada: ISKO-España, 1999, p. 133-140.

Luhn, H.P. (1957) "A statistical approach to mecanized encoding and searching of literary information". *IBM journal of research and developement*, 1957, 1(4), p. 309-317.

MacClure, C. (1994) "User-based data collection techniques and strategies for evaluating networked information services". *Library Trends*, 1994, 42(4), p. 591-607.

MacEwan, A. (2000) "Crossing language barriers in Europe: linking LCSH to other subject heading languages". *Cataloging & Classification Quarterly*, 2000, 29(1/2), p. 199-207.

MacGarry, D.; Svenonius, E. (1991) "More on improved browsable displays for online subject access". *Information Technology & Libraries*, 1991, 10, p. 185-191.

MacIlwaine, I. (1997) "The Universal Decimal Classification: some factors concerning its origins, development and influence." *Journal of the American Society for Information Science*, 1997, 48(4), p. 331-339.

MacKiernan, G. (2002) *Proyect Aristotle(sm)- Automated Categorization of Web Resources* [en línea]. <<http://www.public.iastate.edu/~CYBERSTACKS/Aristotle.htm>> [Consulta: 2 de abril de 2003]

Mandatory Data Elements for Internationally Shared Resource Authority records [en línea]. Report of the IFLA UBCIM Working Group on Minimal Level Authority Records and ISADN, chair Barbara B. Tillet, Françoise Bourdon, Alan Danskin, Andrew Mac Ewan, Eva Murtomaa, Mirna Willer. International Federation of Library Associations and Institutions, Universal Bibliographic

- Control and International MARC Programme, 1998.
<<http://www.ifla.org/VI/3/p1996-2/mlar.htm>> [Consulta 25 de marzo de 2003]
- Maniez, J. (1997) "Database merging and the compatibility of indexing languages".
Knowledge Organization, 1997, 24(4), p.213-224.
- Mano González, M. de la (1994) "El impacto del OPAC en el universo bibliotecario español: ¿un reto para la formación de usuarios?" En: Jornadas Españolas de Documentación Automatizada (4ª. 1994. Gijón). *Los Profesionales ante el reto del siglo XXI: integración y calidad: Documat'94 (Gijón, 6,7 y 8 de octubre 1994)*. Oviedo: Universidad de Oviedo, 1994, p. 565-571
- MARC 21 Concise Format for Classification Data* [en línea]. Library of Congress Network Development and MARC Standards Office. 2002 concise ed.
<<http://www.loc.gov/marc/classification/eccdhome.html>> [Consulta 3 de marzo de 2003]
- Marchionini, G. (1997) *Information seeking in electronic environments*. 2nd ed. New York: Cambridge University Press, 1997.
- Marchionini, G.; Komlodi, A. (1998) "Design of interfaces for information seeking". En: *Annual Review of Information Science and Technology*. M. E. Williams (ed.) Medford, NJ: Information Today, 1998, 33, p. 89-129.
- Marcos Mora, M. C. (1999) "Diversos modos de acceso a catálogos en línea a través de Internet". *El profesional de la información*, 1999, 8(1/2), p. 7-14.
- Markey, K
V.a Drabenstott, K. M.

- Markey, K. (1984). *Subject searching in library catalogs: before and after the introduction of online catalogs*. Dublin, Ohio, OCLC, 1984
- Markey, K. (1987) "Online catalog users and subject authority information". En: *Authority Control Symposium: Papers presented during the 14th annual ARLIS/NA conference, New York, Feb. 10*. K. Muller (Ed.) Tucson, AZ: Art Libraries Society of North America, 1987, p. 29-46
- Markey, K. (1987a) "Searching and browsing the Dewey Decimal Classification in a online catalog". *Cataloging & Classification Quarterly*, 1987, 7(3), p. 37-68.
- Markey, K. (1989). "Subject searching strategies for online catalogues through the Dewey Decimal Classification". En: *The online catalogue: development and directions*. C. R. Hildreth. London, The Library Association, 1989, p. 61-83.
- Markey, K.; Demeyer A. (1986) *Dewey Decimal Classification Online Project: evaluation of a library schedule and index integrated into the subject searching capabilities of an online catalog: final report to the Council of Library Resources*. Dublin, OH: OCLC Online Computer Library Center, 1986.
- Martínez Arellano, F. F. (1999) "Subject searching in online catalogs including spanish and english material". *Cataloging & Classification Quarterly*, 1999, 28(2), p. 45-56.
- Martínez, A. M.; Pichinini, M. (1996) "Catálogos en línea, usuarios y acceso por materia: problemas y recomendaciones". *Referencias*, 1996, 2(2), p. 8-13.
- Martínez, A. M.; Pichinini, M. (1998) *La búsqueda por materia y la descripción de contenido en el catálogo en línea*. La Plata : Universidad Nacional de La Plata, Facultad de Humanidades y Ciencias de la Educación, 1998.

Massicotte, M. (1986) "Improved browsable displays for online subject access". *Information Technology & Libraries*, 1986, 7, p. 373-380.

Matthews, J.; et al. (1983). *Using online catalogs: a nationwide survey, a report of a study sponsored by the Council on Library Resources*. New York: Neal-Schuman, 1983

Menéndez Llana, J. M. (2001) "El uso del OPAC en bibliotecas públicas asturianas", *Boletín de la Asociación Asturiana de Bibliotecarios, Archiveros, Documentalistas y Museólogos*, 2001, enero-junio, p. 13-16.

Michalak, T. J. (1990) "An experiment in enhancing catalog records at Carnegie Mellon University". *Library Hi Technology*, 1990, 8(3), p. 33-41.

Michel, Dee. (1997) *Taxonomy of subject relationships* [en línea]. <<http://www.ala.org/alcts/organization/ccs/sac/msrscu2.pdf>> [Consulta: 23 de abril de 2003]

Milstead, J. L. (1999) "Cross-file searching: how vendors help and don't help improve compatibility". *Searcher*, 1999, 7, p. 44-55.

Mira: evaluations frameworks for interactive multimedia information retrieval applications [en línea]. <<http://www.dcs.gla.ac.uk/mira.>> [Consulta 25 de febrero de 2003]

Mischo, W. H.; Lee, J. (1987) "End-user searching of bibliographic databases". En: *Annual Review of Information Science and Technology*. M. E. Williams (ed.) Amsterdam: Elsevier, 1987, v. 22, p. 227-263.

- Moreiro González, J.A. (2002) "Aplicaciones al análisis automático del contenido provenientes de la teoría matemática de la información". *Anales de documentación*, 2002, 5, p. 273-286.
- Moreiro González, J.A. ; Méndez Rodríguez, E.M. "Lenguaje Natural e Indización automatizada." *Ciencias de la Información*, 1999, 30 (3), p.11-24
- Moreiro González, J. A., et al. (2002) "Mapas conceptuales, topic maps y tesauros". En: *Jornadas de Tratamiento y Recuperación de Información (1ª. 2002. Valencia). Actas de las I Jornadas de Tratamiento y Recuperación de Información (JOTRI): Facultad de Informática Valencia, 4 y 5 de julio de 2002.* E. Sanchis, L. Moreno, I. Gil (Eds.) Valencia: Universidad Politécnica de Valencia, 2002, p. 79-86.
- Moscoso, P. (1998) "Análisis y evaluación de los catálogos públicos de acceso en entorno Web". *Revista española de documentación científica*, 1998, 21(1), p. 409-416.
- Moscoso, P.; Ortiz-Repiso Jiménez, V.; Ríos, Y. (1998) "Telenet versus Web en el acceso a la información catalográfica : estudio comparativo del caso español". En: *Jornadas Españolas de Documentación (6ª. 1998. Valencia). Los sistemas de información al servicio de la sociedad: actas de las Jornadas.* Valencia: FESABID, 1998, p. 641-648.
- Moya Anegón, F. de (1995) *Los sistemas de gestión bibliotecaria: estructuras de datos y recuperación de información.* Madrid: ANABAD, 1995.
- Moya Anegón, F. de (2000) "La investigación española en recuperación de información (RI): análisis bibliométrico". *Revista de investigación iberoamericana en ciencia de la información y documentación*, 2000, 1(1), p. 117-123.

- Moya Anegón, F. de ; Herrero Solana, V. (1999) "Investigaciones en curso sobre interfaces gráficas en dos y tres dimensiones para el acceso a la documentación electrónica" [en línea]. *Cuadernos de documentación multimedia*, 1999, 8. <<http://www.ucm.es/info/multidoc/multidoc/revista/num8/moya.html>> [Consulta: 28 de abril de 2003]
- Moya Anegón, F. de; y Jiménez Contreras, E. (1999) "Autores españoles más citados en Biblioteconomía y Documentación". *El profesional de la información*, 1999, 8(5), p. 8-9.
- Muñoz García, A. (1993) "Técnicas de recuperación de información en bases de datos documentales". *Cuadernos de ADAB*, 1993, 1(2), p. 389-403.
- Mustafa el Hadi, W. (2002) "Terminology & Information retrieval: new tools for new needs". En: *Challenges in Knowledge representation and organization for the 21st Century: integration knowledge across boundaries: proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain*. M. J. López-Huertas (ed.) Würzburg: Ergon, 2002, p. 149-157.
- Nahl, D. (1998) "The user-centered revolution". En: *Encyclopedia of Library and Information Science*. A. Kent and C. M. Hall. New York: Marcel Dekker, 1998. v.62, supl. 25, p. 313-371.
- Nelson, M. J. (1992) "The desing of an hypertext interface for information retrieval". *The canadian journal of information science*, 1992, 16(2), p. 1-12.
- Nicholson, D.; Neill, S. (2001) "Interperability in subject terminologies: the HILT project". *New review of Information networking*, 2001, 7, p. 147-157.

Norgard, B. A.; et al. (1993) "The online catalog: from technical services to access service [en línea] ". *Advances in Librarianship*, 1993, 17, p. 111-148. <<http://bliss.sims.berkeley.edu/papers/advances/advances.txt>> [Consulta: 28 de abril de 2003]

Norma UNE 50-106-90. *Directrices para el establecimiento y desarrollo de tesauros monolingües*. Madrid: AENOR, 1990

O'Brien, A. (1994) "Online catalogs: enhancements and developments". En M. E. Williams (ed.). *Annual Review of Information Science and Technology*. Medford, NJ: Learned Information, 1994, 29, p. 219-242.

Olson, H. A.; Boll, J. J. (2001) *Subject analysis in online catalogs*. Englewood, Col.: Libraries Unlimited, 2001.

Olson, T. (2002) "The integration of information languages and interoperability" presented at the Program *Real world steps to interoperability in libraries* sponsored by the LITA/ALCTS Authority Control in the Online Environment Interest Group, American Library Association Annual Conference, June 16, 2002, Atlanta, Georgia. <<http://www.lita.org/igs/Acig/2002authcontrol.pdf>> [Consulta: 27 de septiembre de 2002]

Olvera Lobo, M. D. (1999) "Evaluación de sistemas de recuperación de información: aproximaciones y nuevas tendencias". *El profesional de la información*, 1999, 8(11), p. 4-14.

Ortego de Lorenzo-Cáceres, P.; Bonal Zazo, J.L. (1996). "Indicadores para el control de calidad de bases de datos bibliográficas españolas". En: Jornadas Españolas de Documentación Automatizada (5ª. 1996. Cáceres). *Sistemas de información: balance de 12 años de jornadas y perspectivas de futuro: actas de*

- las V Jornadas Españolas de Documentación Automatizada*. Cáceres:, Universidad de Extremadura: ABADMEX, 1996, p. 503-512
- Ortiz-Repiso, V. (1998) "Nuevas perspectivas para la catalogación: metadatos versus MARC". *Revista española de documentación científica*, 1998, 22(2), p. 198-219.
- Ortiz-Repiso, V.; Moscoso, P. (1999). "Web-based OPAC's: between tradition and innovation". *Information Technology and Libraries*, 1999, 18(2), <http://www.lita.org/ital/1802_moscoso.html> [Consulta: 20 de marzo de 2003]
- OVER, P. (2001) "The TREC interactive track: an annotated bibliography". *Information Processing & Management*, 2001, 37(3), p. 369-381.
- Paepcke, S. B. (1996). "Using distributed objects for digital library interoperability". *IEEE Computer*, 1996, 29(5), p. 61-69.
- Park, T. K. (1994) "Toward a theory of user based relevance: a call for a new paradigm of inquiry". *Journal of the American Society for Information Science*, 1994, 45(3), p. 135-141.
- Patton, G. E. (2003) "FRANAR: A conceptual model for authority data [en línea] ". En: *Authority control: definition and international experiences: International conference : Florence, Italy, February 10-12 2003*. <<http://www.unifi.it/biblioteche/ac/en/home.htm>> [Consulta: 22 de marzo de 2003]
- Peis, E. (2000) "Tablas de contenido de monografías con carácter colectivo y enriquecimiento de registros bibliográficos". *Boletín de la Asociación Andaluza de Bibliotecarios*, 2000, 59(1), p. 23-34.

Peis, E.; Fernández-Molina, J.C. (1994) "Evaluación de la recuperación de información por materias en los OPACs". En: Jornadas Españolas de Documentación Automatizada (4ª. 1994. Gijón). *Los profesionales ante el reto del siglo XXI: integración y calidad: actas, IV Jornadas Españolas de Documentación Automatizada, DOCUMAT 94 (Gijón, 6, 7 y 8 de octubre 1994)*. Oviedo: Universidad, 1994, p. 245-251.

Peis, E., Fernández Molina, J. C. (1998) "Enrichment of bibliographic records of online catalogs through OCR and SGML technology". *Information Technology and Libraries*, 1998, 17(3), p. 161-172.

Pérez-Carballo, J.; Strzalkowski, T. (2000) "Natural language information retrieval: progress report". *Information Processing & Managenet*, 2000, 36(2), p. 155-178.

Pérez Díez, A. V. (1996) *Perfil y nivel de satisfacción de los usuarios del OPAC de una biblioteca universitaria*. Madrid: CINDOC: FESABID, 1996.

Peters, T. A. (1989). "When smart people fail: an analysis of the transaction log of an online public access catalog". *Journal of Academic Librarianship*, 1989, 15(5), p. 267-273.

Peters, T. A. (1991) *The online catalog: a critical examination of public use*. Jefferson, NC: McFarland, 1991.

Peters, T. A. (1993) "The history and development of transaction log analysis". *Library Hi Tech*, 1993, 11(2), p. 44-66.

Peters, T. A.; Kurth, M. (1991) "Controlled and uncontrolled vocabulary subject searching in an academic library online catalog". *Information Technology and Libraries*, 1991, 10(3), p. 201-211.

Pettigrew, K. E.; Fidel, R.; Bruce, H. (2001) "Conceptual frameworks in information behavior". En: *Annual Review of Information Science and Technology*. M. E. Williams (ed.) Medford, NJ: Information Today, 2001, 35, p. 43-77.

Plaunt, C.; Norgard, B. A. (1997) *A association based method for automatic indexing with a controlled vocabulary* [en línea]. <<http://www.sims.berkeley.edu/research/metadata/assoc/assoc.html>> [Consulta: 2 de abril de 2003]

Poo, D. C. C.; Khoo, C. (1997) "Subject searching in online catalog systems". En: *Encyclopedia of Library and Information Science*. A. Kent (Ed.) New York: Marcel Dekker, 1997, 60 (23), p. 324-341.

Porter, M.; Galpin, V. (1988) "Relevance feedback in public access catalogue for a research library: MUSCAT at the Scott Polar Research Institute". *Program*, 1988, 22(1), p. 1-20.

Porter, M.F. "An algorithm for suffix stripping". *Program*, 1980, 14(3), p. 130-137.

Posey, E. D.; Erdmann, C. A. (1986) "An online UNIX-Based Engineering Library Catalog: Purdue University Engineering Library". *Science & Technology Libraries*, 1986, 32(6:4), p. 31-43.

Potter, W. G. (1989) "Expanding the online catalog". *Information Technology and Libraries*, 1989, 8(2), p. 99-104.

Principles underlying subject headings (1999). Working Group on Principles Underlying Subject Heading Languages. Lopes, M.I., Beall, J. (eds.). Saur: München, 1999.

Pulis, N.; Ludy L. (1988). "Subject searching in an online catalog with authority control". *College and Research Libraries*, 1988, 49(6), p. 523-533.

Qiu, L. (1993) "Analytical searching vs. browsing in hypertext information retrieval systems". *The Canadian Journal of Information and Library Science*, 1993, 18, p. 1-13.

RAMEAU [en línea]: *Répertoire d'autorité matière encyclopédique et alphabétique unifié*. <<http://rameau.bnf.fr>> [Consulta: 10 de marzo de 2003]

Randall, B. N. (1999) "Spelling errors in the database: shadow or substance?". *Library Resources & Technical Services*, 1999, 43(3), p. 161-169.

Ranganathan, S. R. (1933) *Colon Classification*. Madras: Madras Library Association, 1933

Ranganathan, S. R. (1951) *Classified catalogue with additional rules for dictionary catalogue code*. London: Library Association, 1951

Ray, K. L.; Lang, M. S. (1997) "Analyzing search styles of patrons and staff: a replicative study of two university libraries". *Library Resources & Technical Services*, 1997, 41(3), p. 219-235.

RENARDUS [en línea]: *final report*
<http://www.renardus.org/about_us/final_report_def.doc> [Consulta: 2 de abril de 2003]

Répertoire de vedettes-matière [en línea]. Québec : Bibliothèque de l'Université Laval, 2002- . <<http://www.nlc-bnc.ca/rvmweb/index-f.htm>> [Consulta: 10 de marzo de 2003]

Requeriments for a format for classification data [en línea] (1996). IFLA Section on Classification and Indexing and Information Technology Joint Working Group on a Classification Format. <<http://www.ifla.org/VII/s29/projects/rep0796.htm>> [Consulta: 25 de febrero de 2003]

Riesthuis, G.J.A. (1998) "Decomposition of UDC-numbers and the text of the UDC Master Reference File". En: *Structures and relations in knowledge organization: 5th International Isko Conference, Lille, 25-29 August 1998*. W. M. El Hadi, J. Maniez, & S. A. Pollit (Eds.) Würzburg: Ergon, 1998, p. 221-228.

Riesthuis, G.J.A. (2001) "Information languages and multilingual subject access". En: *Subject retrieval in a networked environment: papers presented at an IFLA Satellite Meeting sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology: OCLC, Dublin, Ohio, USA, 14-16 August 2001*. Dublin, OH: OCLC, 2001, p. 231-239.

Riesthuis, G.J.A. ; Zumer, M. (2002) "The functional requirements for bibliographic records and knowledge organization". En: *Challenges in Knowledge representation and organization for the 21st Century: integration knowledge across boundaries: proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain*. M. J. López-Huertas (ed.) Würzburg: Ergon, 2002, p. 165-172.

Ríos García, Y. (1991) "Catálogos en línea de acceso público: selección bibliográfica". *Revista española de documentación científica*, 1991, 14(2), p. 137-145

Robertson, S. E. (1997)"Overview of the OKAPI Projects". *Journal of Documentation*, 1997, 53(1), p. 3-7.

Robertson, S. E.; Hancock-Beaulieu, M. M. (1992). "On the evaluation of IR systems". *Information Processing & Management*, 1992, 28(4), p. 457-466.

Robertson, S. E.; Walker, S.; Beaulieu, M. (2000) "Experimentation as a way of life: Okapi at TREC". *Information Processing & Management*, 2000, 36(1), p.95-108.

Rodríguez Bravo, B.; Alvite Díez, M. L.(2001) "Interfaces web de OPACs en las bibliotecas universitarias españolas [en línea] ". En: Jornades Catalanes de Documentación (8ª. 2001. Barcelona). *Catalunya en la societat de la informació: realitats i reptes, Barcelona 15-16 de novembre de 2001*. Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya: Societat Catalana de Documentació i Informació, 2001. <http://www.cobdc.org/09jornades/8es_Jornades/comunicacions/04_06comunic.pdf> [Consulta: 25 de febrero de 2003]

Roger, D.; Lavandier, J.; Kolmayer, E. (2002) *Navigation et interfaces* [en línea]: *cartes conceptuelles et autres outils*. Réalisée pour PARINFO (Programme d'Aide à la Recherche en INFORMATION), dans le cadre d'une recherche menée a l'ENSSIB. <<http://www.wnssib.fr/bibliotheque/documents/travaux/kolmayer/kolm-navig0.html>> [Consulta: 2 de abril de 2003]

Roose, T.; Lester, M. A.(1988) "Online catalogs: making them better reference tools". *Library Journal*, 1988, 113(20), p. 76-77.

Rosa Piñero, A. de la; Senso Ruiz, J. A.; Eito Brun, J. A. (1998). "Norma Z39.50, actualidad, posibilidades. ¿Es necesario un cambio de actitud?". *Revista española de documentación científica*, 1998, 21(4), p. 121-141

- Rowley, J. (1994) "The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research". *Journal of Information Science*, 1994, 20(2), p. 108-119.
- Saadani, L.; Bertrand-Gastaldy, S. (2003) "Cartes conceptuelles et thésaurus: essai de comparaison entre deux modèles de représentation issus de différentes traditions disciplinaires" [en línea]. En: Association canadienne des sciences de l'information. Congrès annuel (28. 2000. Motreal). *ACSI 2000: Les dimensions d'une science de l'information globale*. <<http://www.slis.ualberta.ca/cais2000/saadani.htm>> [Consulta: 2 de abril de 2003]
- Salton, G. (1986). "On the use of knowledge-based processing in automatic text retrieval". En: *ASIS '86. Proceedings of the 49th ASIS Annual Meeting 1986 : Chicago September 28 - October 2*. J. M. Hurd. Medford, NJ, Learned Information, 1986, 23, p. 277-287.
- Salton, G.; Buckley, C.; Fox, E. A. (1983) "Automatic query formulations in information retrieval". *Journal of the American Society for Information Science*, 1983, 34(4), p. 262-280.
- San Segundo Manuel, R. (1996) *Sistemas de organización del conocimiento: la organización del conocimiento en las bibliotecas españolas*. Madrid: Universidad Carlos III de Madrid, Boletín Oficial del Estado, 1996.
- San Segundo Manuel, R. (1999) "Futura implantación de los distintos sistemas de clasificación en las bibliotecas del Estado español" [en línea]. En: Jornades Catalanes de Documentació (7^a. 1999. Barcelona). *Les biblioteques i els centres de documentació al segle XXI: peça clau de la societat de la informació: Barcelona 4,5 i 6 de novembre de 1999*.

<<http://www.cobdc.org/09jornades/7es/home-cat.htm>.> [Consulta: 6 de marzo de 2003]

San Segundo Manuel, R. (1999a) "Indización en cadena y su aplicación práctica". En: *Sociedad Internacional para la Organización del Conocimiento. Capítulo Español. Congreso (4º. 1999. Granada) La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información: actas del IV Congreso ISKO-España EOCONSID '99, 22-24 de abril de 1999, Granada.* . López Huertas, M.J., Fernández Molina, J. C. (eds.). Granada: ISKO-España, 1999, p. 53-59.

Saracevic, T. (1996) "Modeling interaction in Information Retrieval (IR): a review and proposal". En: *Global complexity: information, chaos and control: proceedings of the American Society for Information Science (ASIS) 59th Annual Meeting.* S. Hardin. Medford, Information Today, 1996, 33, p. 3-9.

Saracevic, T. (1996a) "Relevance reconsidered '96". En: *CoLIS 2: 2nd International Conference on Conceptions of Library and Information Science: integration in perspective: 1996 October 13-16 Copenhagen, Denmark.* P. Ingwersen & N. O. Pors (Eds.) Copenhagen: Royal School of Librarianship, 1996, p. 201-218.

Saracevic, T.; Kantor, P. (1988). "A study of information seeking and retrieving. II: Users, questions, and effectiveness". *Journal of the American Society for Information Science*, 1988, 39(2), p. 177-196.

Saracevic, T.; Kantor, P. (1988). "A study of information seeking and retrieving. III: Searches, searches and overlap". *Journal of the American Society for Information Science*, 1988, 39(2), p. 197-216.

Saracevic, T.; et al. (1988) "A study of information seeking and retrieving. I: Background and methodology". *Journal of the American Society for Information Science*, 1988, 39, p. 161-176.

Savage-Knepshield, P. A.; Belkin, N. J. (1999) "Interactions in information retrieval: trends over time". *Journal of the American Society for Information Science*, 1999, 50(12), p. 1067-1082.

Savolainen, R. (1993) "The sense-making theory: reviewing the interest of a user-centered approach to information seeking and use". *Information Processing & Management*, 1993, 29(1), p. 13-28.

Schamber, L. (1994) "Relevance and information behavior". En M. E. Williams (ed.). *Annual Review of Information Science and Technology*. Medford: Information Today, 1994, 29, p. 3-48.

Schatz, B. R. (1999). "Federating search of scientific literature". *IEEE Computer*, 1999, 32(2), p. 51-59.

Search support for unfamiliar metadata vocabularies [en línea]. School of Information Management Systems, UC Berkeley. <<http://www.sims.berkeley.edu/research/metadata>> [Consulta: 2 de abril de 2003]

Seguí i Palou, R.; Vall, A. (1999) "La consulta por materias en el catálogo automatizado de las bibliotecas públicas de Catalunya". En: Sociedad Internacional para la Organización del Conocimiento. Capítulo Español. Congreso (4º. 1999. Granada). *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información: actas del IV Congreso ISKO-España EOCONSID '99, 22-24 de*

abril de 1999, Granada. . López Huertas, M.J., Fernández Molina, J. C. (eds.). Granada: ISKO-España, 1999, p. 347-352.

Seymour, S. (1991) "Online public access catalog user studies: a review of research methodologies march 1986-november 1989". *Library and Information Science Research*, 1991, 13(1), p. 89-102.

Shaw, D. (1995). "Bibliographic database searching by graduate students in language and literature: search strategies, system interfaces, and relevance judgments". *Library & Information Science Research*, 1995, 17(3), p. 327-345.

Shneiderman, B. (1998) *Designin the user interface: strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley, 1998.

Siegfried, S.; Bates, M. J.; Wilde, D. N. (1993) "A profile of end-user searching behavior by humanities scholars: the Getty online searching project report nº 2". *Journal of the American Society for Information Science*, 1993, 44(5), p. 273-277.

Silkes, S. R. (1997) "Online information retrieval in school library media centers: a cognitive viewpoint". *Public and access services quarterly*, 1997, 2(3), p. 25-44.

Slone, D. J. (2000) "Encounters with the OPAC: on-line searching in public libraries". *Journal of the American Society for Information Science*, 2000, 51(8), p. 757-773.

Slype, George van (1991) *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid; Salamanca: Fundación Germán Sánchez Ruipérez; Madrid: Pirámide, 1991.

- Snelson, P. (1993) "Relationships between access and use in information systems: remote access to and browsing of online catalogues". En : *Proceedings of the 56th Annual Meeting of the American Society for Information Science, Columbus, Ohio, 24-28 October 1993*. S. Bonzi (ed.). Medford, New Jersey: Learned Information, 1993, p.73-80.
- Soergel, D. (1994) "Indexing and retrieval performance: the logical evidence". *Journal of the American Society for Information Science*, 1994, 45(8), 589-599.
- Solomon, P. (1993) "Children`s information seeking behaviour: a case analysis of an OPAC". *Journal of the American Society for Information Science*, 1993, 44, p. 245-264.
- Sonnenwald, D. H.; Iivonen, M. (1999) "An integrated human information behavior research framework for information studies". *Library & Information Science Research*, 1999, 21(4), p. 429-457.
- Sparck Jones, K. (1999) "What is the role of NLP in text retrieval?". En: *Natural language information retrieval*. Strzalkowski (Ed.). Dordrecht: Kluwer Academic Publishers, 1999, p. 1-24
- Sparck Jones, K.; Willet, P.(eds.) (1997) *Readings in information retrieval*. San Francisco: Morgan Kaufmann, 1997.
- Spink, A. (1996) "Multiple search sessions model of end-user behavior: an exploratory study". *Journal of the American Society for Information Science*, 1996, 47(8), p. 603-609.
- Spink, A. [et al.] (1998) "Modeling user's successive searches in digital environments [en línea]". *D-Lib Magazine*, 1998. <<http://www.dlib.org/dlib/april98/04spink.html>> [Consulta: 26 de mayo 2000]

- Spink, A.; Greisdorf, H.; Bateman, J. (1998) "From highly relevant to no relevant: examining different regions of relevance". *Information Processing & Management*, 1998, 35(5), p. 599-622.
- Spink, A.; Losee, R. M. (1996) "Feedback in information retrieval". En: *Annual Review of Information Science and Technology*. M. E. Williams (ed.) Medford, N.J: Information Today, 1996, 31, p. 33-78.
- Spink, A.; Saracevic, T. (1997) "Interaction information retrieval: selection and effectiveness of search terms". *Journal of the American Society for Information Science*, 1997, 48(8), p. 741-761.
- Stone, S. (1982) "Humanities scholars: information needs and uses". *Journal of Documentation*, 1982, 38(4), p. 292-313.
- Strzalkowski, T.; et al. (1999) "Evaluating natural language processing techniques in information retrieval". En: *Natural language information retrieval*. T. Strzalkowski (Ed.) : Kluwer Academic Publishers, 1999, p. 113-145.
- Su, L. T. (1992) "Evaluation measures for interactive information retrieval". *Information Processing & Management*, 1992, 28, p. 503-516.
- Su, S. F. (1994) "Dialogue with an OPAC: how visionary was Swanson in 1964?". *Library Quarterly*, 1994, 64, p. 130-161.
- Subject cataloging manual: Subject headings* (1985). Library of Congress, Cataloging Policy and Support Office. Washington, D.C. : The Library, 1985

- Sugar, W. (1995) "User-centered perspective of information retrieval research and analysis methods". En: *Annual Review of Information Science and Technology*. M. E. Willians (ed.). Medford: Information Today, 1995. 30, p. 77-109.
- Sutcliffe, A. G.; Ennis, M.; Watkinson, S. J. (2000) "Empirical studies of end-user information searching". *Journal of the American Society for Information Science*, 2000, 51(13), p. 1211-1231.
- Svenonius, E. (1983) "Use of classification in online retrieval." *Library Resources & Technical Services*, 1983, 27(1), p. 76-80.
- Svenonius, E. (2000) *The intellectual foundation of information organization*. 2nd printing. Cambridge: Massachusetts Institute of Technology, 2001.
- Svenonius, E. (2000a) "LCSH: Semantics, syntax and specificity". *Cataloging & Classification Quarterly*, 2000, 29(1/2), p. 17-30.
- Swigger, K. (1992) "Scholars' access to the documents they cite: social scientists' information sources". *Annual Review of OCLC Research*, July 1991-June 1992, 61-63.
- Tagliacozzo, R.; Kochen, M. (1970) "Information seeking behavior of catalog users". *Information Storage and Retrieval*, 1970, 6, p. 363-381.
- Tartaglia, S. (2003) "Authority control and subject indexing languages [en línea] ". En: *Authority control: definition and international experiences: International conference : Florence, Italy, February 10-12 2003*. <<http://www.unifi.it/biblioteche/ac/en/home.htm>> [Consulta: 21 de marzo de 2003]

Taylor, A. G. (1984) "Authority files in online catalogs: an investigation of their value". *Cataloging & Classification Quarterly*, 1984, 4(3), p. 1-17.

Tenopir, C.; Nahl-Jakobovits, D.; Howard, D.L. (1991) "Strategies and assessments online: novices' experience". *Library and Information Science Research*, 1991, 13(3), p. 237-266.

Text REtrieval Conference [en línea]. <<http://trec.nist.gov>> [Consulta: 25 de febrero de 2003]

Thomas, D. H. (2001) "The effect of interface desing on item selection in an online catalog". *Library Resources & Technical Services*, 2001, 45(1), p. 20-46.

Tibbo, H. R. (1991) "Information Systems, services, and technology for the Humanities". En: *Annual Review of Information Science and Technology*. M. E. Willians (ed.) Medford: Information Today, 1991, 26, p. 38-73.

Tibbo, H. R. (1994) "Indexing for the Humanities". *Journal of the American Society for Information Science*, 1994, 45(8), 607-619.

Tillet, B.B (ed.) (1989). *Authority control in the online environment: considerations and practices*. New York: Harworth Press, 1989

Tillet, B.B. (1996) "Cataloging rules and conceptual models" [en línea]. <<http://www.oclc.org/research/publications/arr/1996/tillet.htm>> [Consulta: 20 de marzo de 2003]

Tillet, B. B. (2000). "Authority control on the Web" [en línea]. En: *Confronting the challenges of networked resources and the web: Bicentennial Conference on Bibliographic Control for the New Millennium*.

<http://lcweb.loc.gov/catdir/bibcontrol/tillet_paper.html> [Consulta: 15 de abril de 2003]

Tillet, B. B.(2003) "Authority control: state of the art and new perspectives" [en línea]. En: *Authority control: definition and international experiences: International conference : Florence, Italy, February 10-12 200.* <<http://www.unifi.it/biblioteche/ac/en/home.htm>> [Consulta: 20 de marzo de 2003]

Tillotson, J. (1995) "Is keyword searching the answer?". *College & Research Libraries*, 1995, 56, p. 199-206.

Tong, R. M.; et al. (1985) "RUBRIC: an environment for full text information retrieval". En: *Proceedings of the 8th International ACM SIGIR Conference*. Montreal: ACM, 1985, p. 243-251.

UNE 50-106-90. Directrices para el establecimiento y desarrollo de tesauros monolingües. Madrid : AENOR, 1990

USMARC format for classification data; including guidelines for content designation. Washington, DC: Library of Congress Cataloging Distribution Service, 1991.

Vakkari, P.; Kuokkanen, M. (1997) "Theory growth in information science: applications of the theory of science to a theory of information seeking". *Journal of Documentation*, 1997, 53(5), p. 497-519.

Vickery, A.; et al. (1987) "A reference and referral system using expert system techniques". *Journal of Documentation*, 1987, 43(1), p. 1-23.

Vickery, B. (1973) *Information systems*. Hamdem, CT: Archon Books, 1973.

Vickery, B.; Vickery, A. (1993) "Online search interface desing". *Journal of Documentation*, 1993, 49, p. 103-187.

Vizine-Goetz, D.; Godby, J.(2001) *Library Classification Schemes and Access to Electronic Collections: Enhancement of the Dewey Decimal Classification with Supplemental Vocabulary* [en línea]. <http://staff.oclc.org/~vizine/sig_cr/sigcr_done_dvg.htm> [Consulta: 23 de abril de 2003]

Vizine-Goetz, D.; Drabenstott, K. M. (1991) "Computer and manual analysis of subject terms entered by online catalog users". En: ASIS '91. Systems understanding people. Proceedings of the 54 Annual Meeting of the American Society for Information Science , Washington, 27-31 October 1991. J.-M. Griffiths (Ed.) Medford: Learned Information, 1991, v. 28, p. 156-161.

Vossen, P. (ed.) (1998). *EurowordNet: A Multilingual database with lexical semantic networks*. Dordrecht: Kluwer Academic, 1998.

Walker, D. E. (ed.). (1971). *Interactive bibliographic search: the user/computer interface* . Montvale, NJ: AFIPS Press, 1971

Walter, D. R. (1987) "The user at the online catalog: a record of unsuccessful keyword searches: another case study". *LASIE*, 1987, 18(3), p. 74-81.

Walker, S.(1989) "The Okapi online catalogue research projects". En: *The online catalogue: development and directions*. C. R. Hildreth. London, The Library Association, 1989, p. 84-106.

- Walker, S.; De Vere, R. (1990) *Improving subject retrieval in online catalogues. 2. Relevance feedback and query expansion*. London: British Library Research and Development, 1990.
- Walker, S.; Jones, R. M. (1987) *Improving subject retrieval in online catalogues. 1. Stemming, automatic spelling correction and cross-reference tables*. London: British Library Research and Development, 1987.
- Wallace, P. M. (1993) "How do patrons search the online catalog when no one's looking? Transaction log analysis and implications for bibliographic instruction and system design." *RQ*, 1993, 33(2), p. 239-252.
- Wang, P.; Hawk, W. B.; Tenopir, C. (2000) "Users' interaction with World Wide Web resources: an exploratory study using a holistic approach". *Information Processing & Management*, 2000, 36(2), p. 229-251.
- Westbrook, L. (1994) "Qualitative research methods: a review of major stages, data analysis techniques, and quality controls". *Library & Information Science Research*, 1994, 16(3), p. 241-254.
- Wiberley, S. E. (1983) "Subject access in the Humanities and the precision of the humanist's vocabulary". *Library Quarterly*, 1983, 53, p. 420-423.
- Wiberly, S. E. (1991) "Habits of humanists: scholarly behavior and new information technologies". *Library Hi-Tech*, 1991, 9(1), p. 17-21
- Wiberley, S. E.; Daugherty, R. A.; Danowski, J. A. (1990) "User persistence in scanning postings of a computer-driven information system: LCS". *Library and Information Science Research*, 1990, 12(4), p. 341-353.

Wiberley, S. E.; Jones, W. G. (1989) "Patterns of information seeking in the Humanities". *College & Research Libraries*, 1989, 50, p. 638-645.

Wiberley, S. E.; et al. (1995) "User persistence in displaying online catalog postings: LUIS". *Library Resources and Technical Services*, 1995, 39(3), p. 247-264.

Wildemuth, B. M.; O'Neill, A. L. (1995) "The "know" in know-item searches: empirical support for user-centered design". *College & Research Libraries*, 1995, 56, p. 265-281.

Wilder, D.; Geenfield, R. (1997) "The Library of Congress Experimental search System". *Library Hi Tech*, 1997, 15(3/4), p. 127-135.

Wilensky, R. (1996). "Toward work-centered digital information services". *IEEE COMPUTER*, 1996, 29(5), p. 37-45.

Wilkes, A.; Nelson, A. (1995) "Subject searching in two online catalogs: authority control vs. non authority control". *Cataloging & Classification Quarterly*, 1995, 20(4), p. 57-79.

Williams, M. D. (1984) "What makes RABBIT run?". *International Journal of Man-Machine Studies*, 1984, 21, p. 333-352.

Williamson, N. J. (1989) "The Library of Congress Classification and the computer: research in progress." *International cataloging & Bibliographic control*, 18(1), 1989, p. 8-12.

Williamson, N. J. (1991) The Library of Congress Classification: preparations for an online system. En: *Tools for knowledge organization and the human interface:*

- proceedings 1st International ISKO-Conferecne Darmstadt, 14-17 August 1990.*
R. Fugmann. Frankfur/Main: Indeks, 1991, p. 202-209.
- Wilson, T. D. (1997) "Information behavior: an interdisciplinary perspective".
Information Processing & Management, 1997, 33, p. 551-572.
- Wilson, T. D. (1999) "Models in information behavior research". *Journal of Documentation*, 1999, 55(3), p. 249-270.
- Wood, A.; et al. (1995). *Hyperspace: web browsing with visualization* [en línea].
<<ftp://ftp.cs.bham.ac.uk/pub/authors/R.J.Hendley/1995b.ps.Z>> [Consulta: 28 de abril de 2003]
- Wool, G. (2000) "Filing and precoordination: how subject headings are displayed in online catalogs and why it matters". *Cataloging & Classification Quarterly*, 2000, 29(1/2), p. 91-106.
- Workshop on Distributed Heterogeneous Knowledge Networks: Boulder, Colorado 1997, NSF Knowledge Networking* [en línea].
<<http://www.sdc.ucar.edu/info/KDI/>> [Consulta: 12 de abril de 2003]
- Wormell, I. (1985) *Subject access project: SAP: improved subject retrieval for monographic publications*, . Lund: Lund University, 1985.
- Wyly, B. J. (1996) "From acces points to materials: a transaction log analysis of access point value for online catalog users". *Library Resources & Technical Services*, 1996, 40(3), p. 211-236.
- Yan, T. W.; et al. (1996) "From user access patterns to dynamic hypertext linking".
Fifth International World Wide Web Conference Hypertext linking, 1996.

<http://www5conf.inria.fr/fich_html/papers/P8/Overview.html> [Consulta: 12 de abril de 2003]

Yee, M. M. (1991) "System desing and cataloging meet the user: user interfaces to Online Public Access Catalogs". *Journal of the American Society for Information Science*, 1991, 42(2), p. 78-98.

Yee, M. M. (1999) "Guidelines for OPAC displays" [en línea]. En: *65th IFLA Council and General Conference, Bangkok, Thailand, August 20-28, 1999*. <<http://www.ifla.org/IV/ifla65/papers/098-131e.htm>> [Consulta: 25 de abril de 2003]

Yee, M.; Layne, S.S. (1998) *Improving online public acces catalogs*. Chicago: American Library Association, 1998.

Yuan, W.; Meadow, C.T. (1999) "A study of the use of variables in information retrieval user studies". *Journal of the American Society for Information Science*, 1999, 50(2), p. 140-150.

Zillhardt, S.; Bourdon, F. (1998) *AUTHOR project* [en línea]: *transnational application of national name authority files: final report* <<http://www.bl.uk/gabriel/projects/pages/cobra/author.pdf>> [Consulta: 23 de abril de 2003]

Zink, S. D. (1991) "Monitoring user search success through transaction log analysis: the WolfPac example". *Reference Services Review*, 1991, 19(1), p. 49-56.

Zwadlo, J. (1997) "We don't need a philosophy of library and information science: we're confused enough already". *Library Quartely*, 1997, 67(2), p. 103-121.

Zweizig, D.; Dervin, B. (1977) "Public library use, users, uses: advances in knowledge of the characteristics and needs of the adult clientele of american public libraries". En: *Advances in Librarianship*. M. J. Voigt and M. K. Harris. New York: Academic Press, 1977. p. 231-255.

Relación de sitios web citados

British Geological Survey <<http://geolib.bgs.ac.uk/>>
Centre for Intelligent Information Retrieval (University of Massachusetts, Amherst)
<<http://ciir.cs.umass.edu/index.html>>
Centre for Interactive Systems Research at the City University
<<http://www soi.city.ac.uk/organisation/is/research/cisr/>>
Cheshire II <<http://elib.cs.berkeley.edu/cheshire>>
CISNE: Catálogo de la biblioteca de la Universidad Complutense de Madrid
<<http://cisne.sim.ucm.es>>
CoBRA+ Working Group on Multilingual Subject Access
<<http://portico.bl.uk/gabriel/projects/pages/cobra/finrap3.html>>
CoBRA+: Computerised Bibliographic Record Actions
<<http://portico.bl.uk/gabriel/cobra/>>
CONNEXION <<http://connexion.oclc.org/>>
Consortium of University Research Libraries (COPAC) <<http://copac.ac.uk/copac/>>
Digital Libraries Initiative <<http://www.si.umich.edu/SantaFe/>>
Dublin Core Metadata Initiative <<http://uk.dublincore.org/>>
FullView <<http://research.oclc.org/Normore/public/OverView/FullView/fv.asp>>
HyperSpace <<http://www.cs.bham.ac.uk/~nsd/>>
Information Society Technologies Programme
<<http://www.cordis.lu/ist/ist-fp5.html>>
Library of Congress OPAC <<http://catalog.loc.gov>>
LyberWorld
<<http://www-cui.darmstadt.gmd.de/visit/Activities/Lyberworld/demos.html>>
Marc in XML <<http://www.loc.gov/marc/marcxml.html>>

MarcXml: MARC 21 XML Schema oficial web site.

<<http://www.loc.gov/standards/marcxml/>>

MARIAN <<http://www.dlib.vt.edu/products/marian.html>>

Metadata Research Program

<<http://metadata.sims.belkeley.edu/ResearchAreas/NaturalLan.html>>

Oasis <<http://www.sims.belkeley.edu/research/projects/oasis/>>

OCLC Research Projects <http://orc.rsch.oclc.org:6464/related_projects.html>

OCLC Research Projects: Barcelona

<<http://research.oclc.org/Normore/public/OverView/Barcelona/barcelona.asp>>

OCLC Research Projects: Information visualization

<<http://research.oclc.org/Normore/public/OverView/index.html>>

Onion Patch(sm): New age public access systems: projects, research, products and services. <<http://www.public.iastate.edu/~CYBERSTACKS/Onion.htm>>

Text Encoding Initiative <<http://www.tei-c.org/>>

The Big Picture(sm)

<<http://www.public.iastate.edu/~CYBERSTACKS/BigPic.htm>>

UKOLN: United Kingdom Office for Library Information Networking (Universty of Bath) <<http://www.ukoln.ac.uk/>>

VIAF: The Virtual International Authority File

<<http://www.oclc.org/research/projects/viaf/index.html>>

W3C Metadata and resource description <<http://www.w3.org/Metadata/>>

WEBSOM <<http://websom.hut.fi/websom>>

WordView3D

<<http://research.oclc.org/Normore/public/OverView/WordView3D/wv3.asp>>

Anexo: Tablas y gráficos

Tabla A.1 Errores tipográficos: distribución por características

Tipo	% Errores	Estructura	% Errores	Palabras	% Errores
650	77,5	Nombre común + Adjetivo	2,9	2	37,3
600	9,4	Nombre común	21	1	26
651	5,3	Nombre propio	17,2	3	16,5
630	4,4	Nombre común + Preposición + Nombre común	8,2	4	12,6
610	3,5	Otras estructura	< 5	5 o más	7,6

Tabla A.2 Búsquedas no válidas por áreas y tipos

	Número	% de búsquedas en el área	Extranjeras	Letras	Incompletas	Signos
Ciencias sociales	17	4,42	8	7	2	0
Exactas-Naturales	21	5,45	13	5	2	1
Humanidades	20	5,19	9	5	4	2
Médicas	16	4,16	4	7	5	0
Opac-Web	25	6,5	16	1	6	2
Total	99(5,2%)		50(2,7%)	25(1,3%)	19(1%)	5(0,2%) ¹

¹ Las diferencias que puedan existir se deben a errores de redondeo.

Tabla A.3 Búsquedas no válidas: diferencias entre áreas. Prueba Chi-cuadrado

	C. Sociales	ExacNat	Humanidades	Médicas	Opac-Web
Ciencias sociales					
Exactas Naturales	.618				
Humanidades	.737	1			
Médicas	1	.501	.609		
Opac-Web	.266	.649	.539	.199	

Tabla A.4 Errores ortográficos por áreas

	C.Sociales	ExactNat	Humanidades	Médicas	Opac-Web
Ortografía incorrecta	2 0,52	2 0,52	3 0,78	2 0,52	

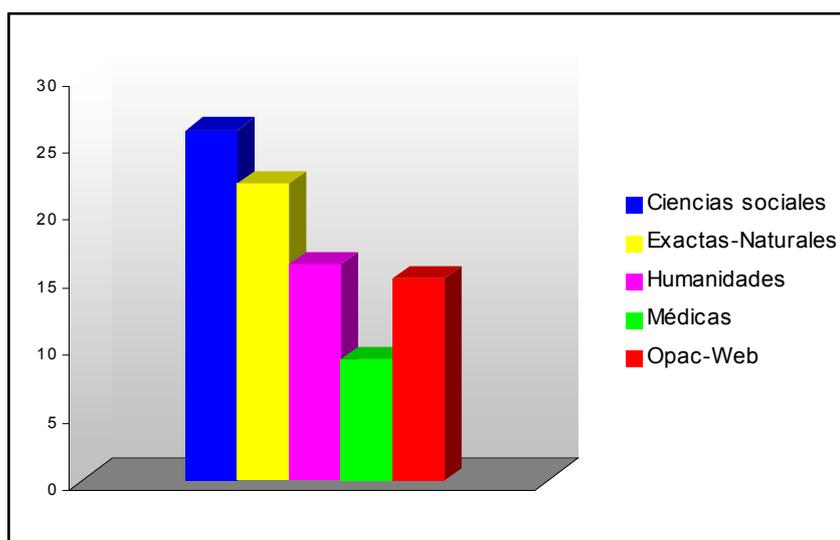


Fig. A.1 Errores tipográficos: distribución por áreas

Tabla A.5 Errores tipográficos: distribución por áreas

	Ciencias sociales			C. Exactas-naturales			Humanidades			Ciencias médicas			Opac Web		
	N° error	N° búsq	%	N° error	N° búsq	%	N° error	N° búsq	%	N° error	N° búsq	%	N° error	N° búsq	%
Sustitución	7	7	1,82	5	5	1,30	5	5	1,30	2	2	0,52	7	7	1,82
Permutación	3	2	0,52	2	2	0,52	2	2	0,52	1	1	0,26	2	1	0,26
Inserción	4	4	1,04	8	7	1,82	2	2	0,52	3	2	0,52	3	3	0,78
Omisión	22	16	4,16	16	11	2,86	10	10	2,60	5	4	1,04	5	5	1,30
Total	36	29	7,53	31	25	6,49	19	19	4,94	11	9	2,34	17	16	4,16

Tabla A.6 Errores tipográficos: diferencias entre áreas. Prueba Chi-cuadrado

	C. Sociales	ExacNat	Humanidades	Médicas	Opac-Web
Ciencias sociales					
Exactas Naturales	.551				
Humanidades	.113	.318			
Médicas	.003	.017	.155		
Opac-Web	.077	.238	.855	.213	

Tabla A.7 Número de palabras: porcentaje total

Palabras	Porcentaje
1	36,90
2	30,53
3	15,22
4	7,47
5	2,77
6	1,41
7	0,35
8	0,09
9	0,06
No válidos	5,20

Tabla A.8 Tipos de búsqueda

Campos Marc	Porcentaje
600 Nombre de persona	8,29
610 Nombre de entidad	2,11
630 Título	2,11
650 Materia	78,37
651 Nombre geográfico	3,92
No válidos	5,20

Tabla A.9 Tipos de búsqueda: número de palabras

	Personas	Entidades	Título	Materias	Geográficos
1 palabra	50,17 %	21,86 %	13,81 %	40,99 %	45,37 %
2 palabras	34,53 %	28,39 %	32,38 %	31,04 %	38,21 %
3 palabras	11,04 %	19,93 %	17,05 %	16,26 %	7,67 %
4 palabras	1,26 %	19,06 %	18,09 %	7,52 %	7,62 %
5 o más	3,01%	10,75 %	18,66 %	4,18 %	2,14 %

Tabla A.10 Número de palabras: distribución por áreas

Palabras	Ciencias sociales		Exactas-Naturales		Humanidades		Médicas		Opac-Web	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
1	136	35,32	186	48,31	143	37,14	181	47,01	135	35,06
2	113	29,35	103	26,75	119	30,91	127	32,99	124	32,21
3	68	17,66	45	11,69	51	13,25	30	7,79	59	15,32
4	32	8,31	18	4,68	27	7,01	20	5,19	29	7,53
5	12	3,12	10	2,60	15	3,90	6	1,56	6	1,56
6	5	1,30	2	0,52	7	1,82	2	0,52	6	1,56
7	2	0,52	0	0	1	0,26	1	0,26	1	0,26
8	0	0	0	0	1	0,26	2	0,52	0	0,00
9	0	0	0	0	1	0,26	0	0	0	0
No válido	17	4,42	21	5,45	20	5,19	16	4,16	25	6,49
Media	2,17		1,82		2,15		1,82		2,08	
Desv. típ.	1,23		1,07		1,34		1,12		1,16	

Tabla A.11 Número de palabras: diferencias entre áreas. ANOVA

	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	Opac-Web
Ciencias sociales					
Exactas Naturales	.000				
Humanidades	.686	.001			
Médicas	.000	.775	.001		
Opac-Web	.186	.007	.351	.017	

Tabla A.12 Tipos de búsqueda: distribución por áreas

Campo	Ciencias sociales		Exactas-Naturales		Humanidades		Médicas		Opac-Web	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
600 Nombre de persona	8	2,08	5	1,30	63	16,36	5	1,30	50	12,99
610 Nombre de entidad	9	2,34	0	0,00	8	2,08	0	0,00	10	2,60
630 Título	5	1,30	0	0,00	24	6,23	1	0,26	2	0,52
650 Materia	335	87,01	347	90,13	243	63,12	362	94,03	283	73,51
651 Nombre geográfico	11	2,86	12	3,12	27	7,01	1	0,26	15	3,90
No válidos	17	4,42	21	5,45	20	5,19	16	4,16	25	6,49

Tabla A.13 Tipos de búsqueda: diferencias entre áreas. Prueba Chi-cuadrado

	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	Opac-Web
Ciencias sociales					
Exactas Naturales	.009				
Humanidades	.000	.000			
Médicas	.001	.023	.000		
Opac-Web	.000	.000	.000	.000	

Tabla A.14 Estructura: diferencias entre áreas. Prueba Chi-cuadrado

	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	Opac-Web
Ciencias sociales					
Exactas Naturales	.000				
Humanidades	.000	.000			
Médicas	.000	.000	.000		
Opac-Web	.015	.000	.032	.000	

Tabla A.15 Coincidencia palabras en el índice

	Palabras en el índice		Palabras en un encabezamiento	
	Sin correcciones	Corregido	Sin correcciones	Corregido
1 palabra	6,95	5,12	21,70	18,75
2 palabras	14,84	13,32	6,80	6,19
3 palabras	6,72	6,41	1,00	0,99
4 palabras	0,97	0,97		
5 palabras	0,10	0,10		
Total	29,59	25,93	29,50	25,93

Tabla A.16 Coincidencia palabras en un encabezamiento

	N° palabras en un encabezamiento			
	0	1	2	3
2 palabras en el enunciado	10,09	54,67		
3 palabras en el enunciado	0,59	26,52	36,82	
4 palabras en el enunciado	2,35	12,29	34,00	46,47
5 palabras en el enunciado		3,57	16,04	44,87
6 palabras en el enunciado	2,35	1,57	9,32	1,31
7 palabras en el enunciado		0,55	1,68	7,35
8 palabras en el enunciado		0,14	0,97	
9 palabras en el enunciado			0,97	

Tabla A.17 Porcentajes coincidencia

	Sin correcciones	Corregido
Exacta	37,86	42,58
Parcial	19,61	20,97
Alguna palabra	29,83	26,00
Ningún término	7,2	4,40
No válidos	5,2	5,20

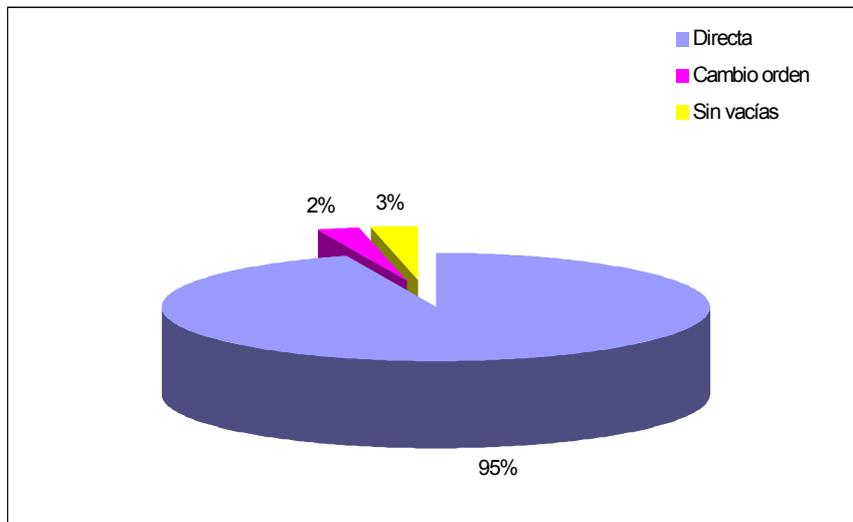


Fig. A.2 Coincidencia exacta: tipos

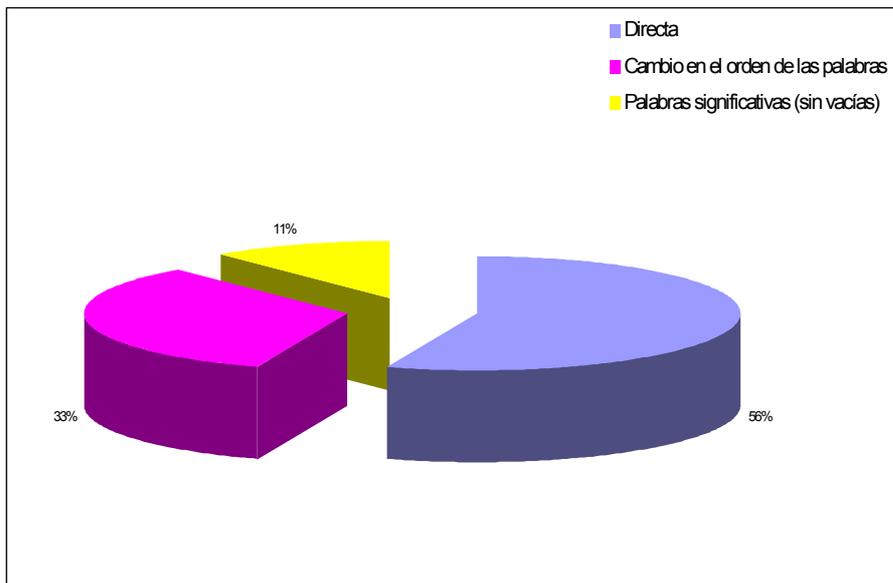


Fig. A.3 Coincidencia parcial: tipos

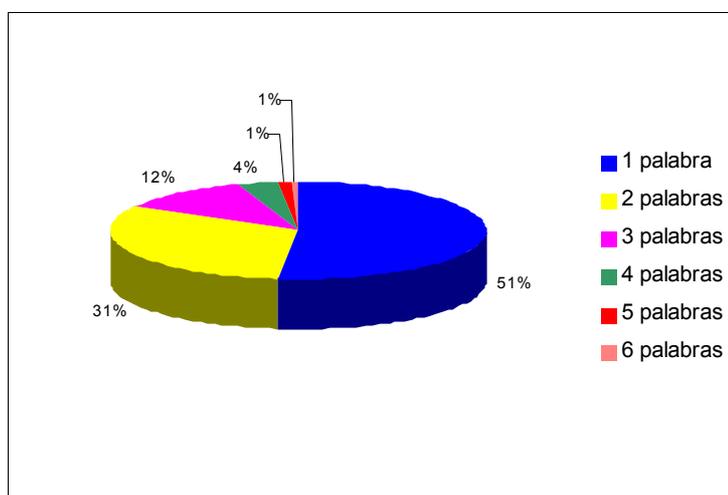


Fig. A.4 Coincidencia exacta: número de palabras

Tabla A.18 N° de palabras/ coincidencia

Palabras	1	2	3	4	5	6	7	8	9
Coincidencia exacta	59,39	43,72	32,66	20,70	16,22	19,89	20,01		
Coincidencia parcial	28,04	20,48	19,55	12,87	11,53	10,13			
	87,44	64,20	52,21	33,58	27,74	30,02	20,01		

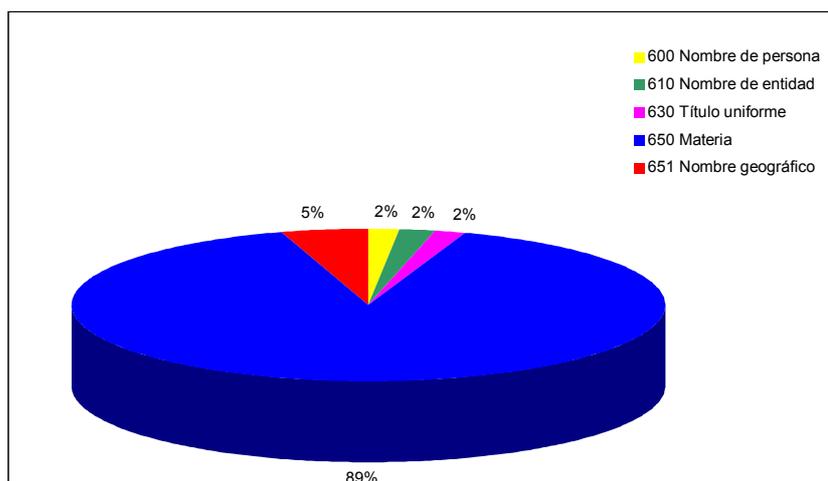


Fig. A.5 Coincidencia exacta: tipo de búsqueda

Tabla A.19 Tipo de búsqueda/Coincidencia exacta y parcial

	600	610	630	650	651
Exacta	8,52	36,32	35,41	48,93	51,53
Parcial	71,68	21,15	14,60	16,40	37,00
Total	80,20	57,47	50,00	65,32	88,53

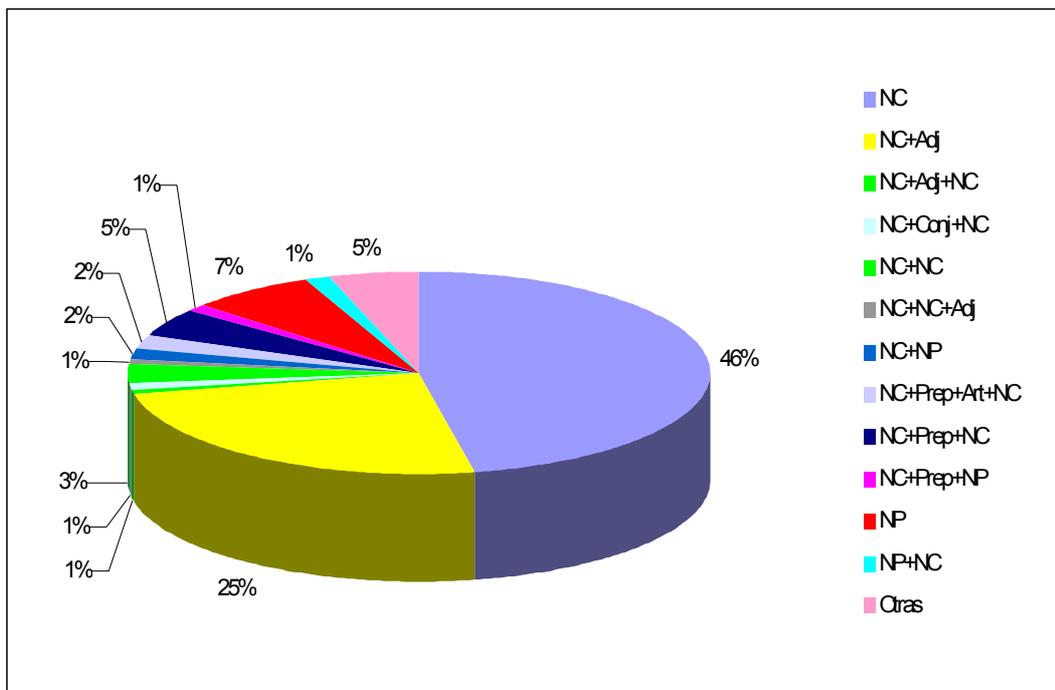


Fig. A.6 Coincidencia exacta: estructura

Tabla A.20 Estructura/Tipo de coincidencia

	NC	NC+Adj	NC+Adj+Adj	NC+Adj+NC	NC+Conj+NC	NC+NC	NC+NC+Adj	NC+NP
Exacta	68,83	50,57	10,78	49,48	26,46	42,56	19,91	37,70
Parcial	19,21	10,76		12,11	11,60	18,87	30,24	16,84
Total	88,04	61,32	10,78	61,59	38,06	61,44	50,15	54,54

Tabla A.20. Estructura/Coincidencia

	NC+Prep+Art+NC	NC+Prep+NC	NC+Prep+NC+Adj	NC+Prep+NP	NP	NP+NC	Otras
Exacta	35,13	43,63	8,99	27,76	21,81	38,74	20,70
Parcial	12,99	16,25		36,42	56,81	37,93	18,01
Total	48,12	59,88	8,99	64,18	78,62	76,67	38,72

Tabla A.21 Ciencias sociales: coincidencia de las expresiones con el índice

Ciencias sociales					
		Sin corrección		Corregidos	
Coincidencia exacta					
Orden directo		119 30,91		137 35,58	
<i>Directa</i>		103 26,75		117 30,39	
<i>Véase</i>		16 4,156		20 5,19	
Cambio en el orden de las palabras		4 1,04		5 1,30	
<i>Subcampos</i>		4 1,039		5 1,30	
<i>Palabras</i>					
Palabras significativas		8 2,08		11 2,86	
<i>Mismo orden</i>		1 0,26		2 0,52	
<i>Cambio de orden</i>					
<i>Subcampos</i>		6 1,558		8 2,08	
<i>Palabras</i>		1 0,26		1 0,26	
Total		131 34,03		153 39,74	
Coincidencia parcial					
Orden directo		26 6,75		28 7,27	
<i>Índice</i>		26 6,75		28 7,27	
<i>Véase</i>					
Cambio en el orden de las palabras		13 3,38		15 3,9	
<i>Subcampos</i>		12 3,12		14 3,64	
<i>Palabras</i>		1 0,26		1 0,26	
Palabras significativas (sin vacías)		8 2,08		8 2,08	
<i>Mismo orden</i>		2 0,52		2 0,52	
<i>Cambio de orden</i>					
<i>Subcampos</i>		3 0,78		3 0,78	
<i>Palabras</i>		3 0,78		3 0,78	
Total		47 12,21		51 13,2	
Coincidencia palabras		34,94		31,57	

Tabla A.22 Exactas-Naturales: coincidencia de las expresiones con el índice

Exactas-Naturales					
		Sin corrección		Corregidos	
Coincidencia exacta					
Orden directo		156 40,52		175 45,45	
<i>Directa</i>		141 36,62		157 40,78	
<i>Véase</i>		15 3,90		18 4,68	
Cambio en el orden de las palabras		4 1,04		5 1,30	
<i>Subcampos</i>		3 0,78		4 1,04	
<i>Palabras</i>		1 0,26		1 0,26	
Palabras significativas		6 1,56		8 2,08	
<i>Mismo orden</i>				1 0,26	
<i>Cambio de orden</i>					
<i>Subcampos</i>		6 1,56		7 1,82	
<i>Palabras</i>					
Total		166 43,12		188 48,83	
				5,71	
Coincidencia parcial					
Orden directo		37 9,61		37 9,61	
<i>Indice</i>		37 9,61		37 9,61	
<i>Véase</i>					
Cambio en el orden de las palabras		24 6,23		27 7,01	
<i>Subcampos</i>		11 2,86		11 2,86	
<i>Palabras</i>		13 3,38		16 4,16	
Palabras significativas (sin vacías)		5 1,30		8 2,08	
<i>Mismo orden</i>		1 0,26		1 0,26	
<i>Cambio de orden</i>					
<i>Subcampos</i>		1 0,26		3 0,78	
<i>Palabras</i>		3 0,78		4 1,04	
Total		66 17,14		72 18,70	
Coincidencia palabras		23,12		20	

Tabla A.23 Humanidades: coincidencia de las expresiones con el índice

Humanidades				
	Sin corrección		Corregidos	
Coincidencia exacta				
Orden directo			116	30,13
				124
				32,21
<i>Directa</i>	102	26,49		110 28,57
<i>Véase</i>	14	3,64		14 3,64
Cambio en el orden de las palabras			3	0,78
				3
				0,78
<i>Subcampos</i>	3	0,78		3 0,78
<i>Palabras</i>				
Palabras significativas			9	2,34
				12
				3,12
<i>Mismo orden</i>	4	1,04		6 1,56
<i>Cambio de orden</i>				
<i>Subcampos</i>	5	1,30		6 1,56
<i>Palabras</i>				
Total			128	33,25
				139
				36,10
Coincidencia parcial				
Orden directo			65	16,88
				69
				17,92
<i>Índice</i>	65	16,88		69 17,92
<i>Véase</i>				
Cambio en el orden de las palabras			33	8,57
				40
				10,39
<i>Subcampos</i>	10	2,60		14 3,64
<i>Palabras</i>	23	5,97		26 6,75
Palabras significativas (sin vacías)			9	2,34
				11
				2,86
<i>Mismo orden</i>	1	0,26		2 0,52
<i>Cambio de orden</i>				
<i>Subcampos</i>				
<i>Palabras</i>	8	2,08		9 2,34
Total			107	27,79
				120
				31,17
Coincidencia palabras			27,09	24,64

Tabla A.24 Ciencias médicas: coincidencia de las expresiones con el índice

Ciencias médicas						
	Sin corrección			Corregidos		
Coincidencia exacta						
Orden directo			192	49,87		205 53,25
<i>Directa</i>	176	45,71			187	48,57
<i>Véase</i>	16	4,16			18	4,68
Cambio en el orden de las palabras			1	0,26		1 0,26
<i>Subcampos</i>	1	0,26			1	0,26
<i>Palabras</i>						
Palabras significativas			7	1,82		9 2,34
<i>Mismo orden</i>	4	1,04			4	1,04
<i>Cambio de orden</i>						
<i>Subcampos</i>	3	0,78			5	1,30
<i>Palabras</i>						
Total			200	51,95		215 55,84
Coincidencia parcial						
Orden directo			19	4,94		17 4,42
<i>Índice</i>	17	4,42			16	4,16
<i>Véase</i>	2				1	0,26
Cambio en el orden de las palabras			13	3,38		19 4,94
<i>Subcampos</i>	4	1,04			4	1,04
<i>Palabras</i>	9	2,34			15	3,90
Palabras significativas (sin vacías)			2	0,52		2 0,52
<i>Mismo orden</i>	1	0,26			1	0,26
<i>Cambio de orden</i>						
<i>Subcampos</i>	1	0,26			1	0,26
<i>Palabras</i>						
Total			34	8,83		38 9,87
Coincidencia palabras			25,47			24,42

Tabla A.25 Opac-Web: coincidencia de las expresiones con el índice

Opac-Web					
	Sin corrección		Corregidos		
Coincidencia exacta					
Orden directo		168	43,64	179	46,49
<i>Directa</i>	162	42,08		172	44,68
<i>Véase</i>	6	1,56		7	1,82
Cambio en el orden de las palabras		3	0,78	5	1,30
<i>Subcampos</i>	3	0,78		4	1,04
<i>Palabras</i>				1	0,26
Palabras significativas		2	0,52	2	0,52
<i>Mismo orden</i>	1	0,26		1	0,26
<i>Cambio de orden</i>					
<i>Subcampos</i>	1	0,26		1	0,26
<i>Palabras</i>					
Total		173	44,94	186	48,31
Coincidencia parcial					
Orden directo		48	12,47	48	12,47
<i>Índice</i>	48	12,47		48	12,47
<i>Véase</i>					
Cambio en el orden de las palabras		18	4,68	21	5,45
<i>Subcampos</i>	4	1,04		6	1,56
<i>Palabras</i>	14	3,64		15	3,90
Palabras significativas (sin vacías)		2	0,52	2	0,52
<i>Mismo orden</i>	1	0,26		1	0,26
<i>Cambio de orden</i>					
<i>Subcampos</i>					
<i>Palabras</i>	1	0,26		1	0,26
Total		68	17,66	71	18,44
Coincidencia palabras		24,34		22,18	

Tabla A.26 Tipo de coincidencia por áreas. Porcentajes

	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	OPAC-Web
Exacta	39,74 %	48,83 %	36,1 %	55,84 %	48,31 %
Parcial	17,1 %	18,7 %	31,17 %	9,87 %	18,44 %
Alguna palabra	31,57 %	20 %	24,64 %	24,42 %	22,18 %
Ningún término	6,6 %	7,01 %	3,11 %	5,97 %	3,64 %
No válidos	4,42 %	5,45 %	5,19 %	4,16 %	6,5 %

Tabla A.27 Coincidencia exacta por áreas. Prueba de Chi cuadrado

	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	OPAC Web
Ciencias sociales					
Exactas-Naturales	.007				
Humanidades	.334	.000			
Médicas	.000	.097	.000		
Opac Web	.017	.829	.001	.041	

Tabla A.28 Coincidencia parcial por áreas. Prueba de Chi cuadrado

	Ciencias sociales	Exactas-Naturales	Humanidades	Médicas	OPAC Web
Ciencias sociales					
Exactas-Naturales	.851				
Humanidades	.000	.000			
Médicas	.001	.002	.000		
Opac Web	.714	.518	.001	.000	

Tabla A.29 Áreas: características/coincidencia exacta. Prueba Chi-cuadrado

	Número de palabras	Tipo	Estructura
Ciencias sociales	.000	.000	.000
Exactas-Naturales	.000	.000	.000
Humanidades	.000	.000	.000
Médicas	.000	.000	.000
Opac Web	.000	.000	.000

Tabla A.30 Áreas: características/coincidencia parcial. Prueba Chi-cuadrado

	Número de palabras	Tipo	Estructura
Ciencias sociales	.096	.202	.000
Exactas-Naturales	.056	.000	.000
Humanidades	.000	.000	.000
Médicas	.636	.044	.021
Opac Web	.010	.000	.000