



**SUBDIRECCIÓN DE SERVICIOS TÉCNICOS Y
ADQUISICIONES**

**SERVICIO DE PROCESO TÉCNICO Y
NORMALIZACIÓN**

**EVALUACIÓN DEL USO DE DESCRIPTORES EN
LOS REGISTROS BIBLIOGRÁFICOS**

**Informe preparado por María Andrés Cuervo, Servicio de Proceso Técnico y Normalización, con la colaboración del Grupo de Trabajo de Materias
Mayo 2006**

ÍNDICE

Página

INTRODUCCIÓN.....	2
TERMINOLOGÍA.....	3
SITUACIÓN ACTUAL EN LA BUC	4
EVALUACIÓN DE LA RECUPERACIÓN ACTUAL POR MATERIAS	5
EVALUACIÓN DEL POSIBLE USO DE DESCRIPTORES.....	6
POSIBILIDAD DEL USO DE DESCRIPTORES LIBRES.....	7
EVALUACIÓN DEL USO DE DESCRIPTORES LIBRES.....	7
CONCLUSIONES.....	9
ANEXOS	
ANEXO 1. NORMAS ISO Y NORMAS UNE.....	12
ANEXO 2. INFORMACION DE GESTIÓN SOBRE BÚSQUEDAS.....	16

INTRODUCCIÓN

El presente informe tiene como objeto evaluar la posibilidad y/o la conveniencia de incluir en los registros bibliográficos formas de descripción del contenido de los documentos como alternativa o complemento de las utilizadas actualmente (Tesauro BUC, palabras clave, CDU)

Para realizar esta evaluación, el Servicio de Proceso Técnico y Normalización ha contado con la colaboración, en primer lugar del Grupo de Trabajo de Materias:

Amanda Cabo (Facultad de Medicina)
Concha Delgado (Facultad de Veterinaria)
María Victoria Romero (Facultad de Ciencias Geológicas)
Gumersindo Villar (Facultad de Informática)
José Chaparro (Facultad de Ciencias de la Información)
Luisa María Payno (Facultad de Derecho)
María Cruz Hermida (Facultad de Geografía e Historia)
Asunción Romero (Facultad de Filología)
Susana Corullón (Facultad de Ciencias Políticas y Sociología)
José Vicente Serrano (Facultad de Ciencias Económicas)

y de 14 centros que han aportado su opinión, su experiencia y sus sugerencias.

TERMINOLOGIA

De acuerdo con las normas expuestas en el Anexo 1, resumimos el sentido de los términos que se van a utilizar:

♦ **Tesaurus:** Vocabulario controlado y dinámico, compuesto por términos que tienen entre ellos relaciones semánticas y genéricas y que se aplica a un dominio particular del conocimiento (Norma ISO-2788) Así pues lo fundamental del tesaurus es, por una parte, ser un lenguaje que utiliza herramientas conceptuales de control para eliminar las ambigüedades semánticas y por otra, ser aplicable a un área específica del conocimiento e incluso hacerlo desde un cierto ángulo en particular..

♦ **Descriptores.** Son los elementos de un tesaurus que constituyen términos indizantes que tienen carácter de preferentes para representar un concepto dado frente a los términos no preferentes o no descriptores que no se utilizan en la indización pero constituyen una entrada alternativa desde la que se puede acceder directamente al término preferente mediante la instrucción USE (equivalencia). Los descriptores son unívocos, la sinonimia se elimina mediante las relaciones de equivalencia ya mencionadas, la polisemia gracias a las relaciones semánticas y jerárquicas y el uso se delimita mediante notas de alcance.

♦ **Encabezamientos de materia.** Constituyen un lenguaje documental de estructura asociativa, al igual que los tesaurus, de los que se suelen distinguir al considerar que éstos son lenguajes postcoordinados (en los que los términos, descriptores y subdivisiones, se combinan en el momento de la recuperación), mientras que los encabezamientos corresponderían a un tipo de lenguaje precoordinado (los términos se combinan en el momento de la indización). Sin embargo la realidad es que en casi todos los casos hay una mayor o menor participación de ambas características.

♦ **Palabras clave.** Son palabras del lenguaje natural utilizadas para la indización mediante la **extracción** automatizada de diversas partes de los documentos primarios o secundarios que se indizan (título, resumen, notas, autor, etc.), frente al sistema de asignación de términos o encabezamientos por el indizador, que es el utilizado en los lenguajes controlados y en los descriptores libres. En estos sistemas, además de eliminarse las palabras vacías mediante listas (stop words), se incorpora algún tipo de técnica para valorar la importancia de esas palabras en la recuperación: frecuencia de aparición, lugar en el que aparecen en el documento, etc.

♦ **Descriptores libres.** Es una forma de indización intermedia entre la extracción y la asignación. Son términos o expresiones elegidos por el indizador sin utilizar ningún lenguaje documental controlado. Coinciden con las palabras clave en que no hay ninguna limitación en el número de expresiones que se pueden utilizar para designar los conceptos y en que estas expresiones no se extraen de ningún vocabulario preestablecido ni están normalizadas. Se diferencian de las palabras clave en que los términos son asignados por el

indizador mediante un proceso intelectual sin necesidad de utilizar palabras que figuren en el documento.

SITUACION ACTUAL EN LA BUC

SISTEMAS DE BÚSQUEDA Y RECUPERACIÓN

■ **El Tesauro BUC y la búsqueda por materias en Cisne.** Elaborado para cubrir la necesidad de disponer de los subencabezamientos y sus notas de uso, ya que el sistema Millennium no permite realizar registros de autoridad de subencabezamiento (aunque el MARC21 sí contempla esta posibilidad), y para proporcionar una herramienta de consulta flexible y de actualización rápida que permite la búsqueda alfabética, por CDU, por palabras clave y mediante el uso de operadores booleanos. Tiene en gran parte estructura de Tesauro, puesto que utiliza salvo excepciones la postcoordinación, la jerarquización, y las relaciones de asociación y equivalencia, aunque su temática no es especializada sino general. Una vez localizados por los usuarios los términos por los que desean buscar información, la utilidad de recuperación está facilitada por la existencia de un enlace al Catálogo Cisne y otro a Z39.50. Por otro lado, el catálogo Cisne cuenta con una recuperación alfabética de materias a partir de un índice específico al que van las etiquetas MARC 6XX y que permite la búsqueda por cualquiera de los subcampos que figuran en dichas etiquetas

■ **Palabras clave.** Millennium realiza la indización de palabras clave en todos los campos que indizan por otros conceptos (autor, materia, título, editor, etc.) más la mayoría de las notas. Los documentos se recuperan en una lista ordenada por frecuencia de aparición de la palabra clave por la que se realiza la búsqueda y ésta se puede refinar por idioma, tipo de material, ubicación, editor y fechas. Esta búsqueda se utiliza no sólo para materias, sino también para autores y títulos y esto hay que tenerlo en cuenta al interpretar las estadísticas.

■ **Descriptores.** La BUC ha definido una de las etiquetas 69X (acceso a materias locales) que ofrece MARC21, como 690 04, para el uso de la Biblioteca Europea que utiliza el Tesauro Eurovoc de las Comunidades Europeas ya que necesitaba este medio para reflejar unas materias muy específicas, nombres de organismos, comisiones, subcomisiones, planes, etc, que, en muchos casos no estaban cubiertas por el Tesauro BUC, ni tendría mucho sentido que lo estuvieran. El contenido de esta etiqueta tiene un índice propio que recupera en su propio scope en el Opac de Millennium.

EVALUACION DE LA RECUPERACION ACTUAL POR MATERIAS

A través de Telnet o de Millennium Web Management Reports se puede acceder a la información de gestión sobre búsquedas en los catálogos públicos de la Universidad Complutense:

Como puede verse en el Anexo 2 se ha tomado a modo de ejemplo la información de gestión de un intervalo de treinta días, tras comprobar en varios intervalos sucesivos que las variaciones porcentuales son mínimas entre ellos.

Hemos centrado la atención en las búsquedas por materia y por palabra clave por ser las más significativas para el tema de este informe, pero en primer lugar hay que hacer una advertencia fundamental: los datos estadísticos sobre ambos tipos de búsqueda no son comparables, puesto que la búsqueda por palabras clave es válida no sólo para las materias sino también para los autores y los títulos y dados los datos que aportan estas estadísticas y por la misma naturaleza de las búsquedas por palabra clave, es imposible saber qué parte de los totales, y por tanto de los porcentajes, se ha dedicado a búsquedas y recuperaciones de uno u otro tipo.

Otro hecho que puede influir en el elevado número de consultas por palabra clave es que la página de inicio de la Web de la biblioteca y la primera opción al seleccionar de ella "Catálogo Cisne", ofrecen por defecto la búsqueda por palabra clave, cosa que, en el primero de los casos ni siquiera se especifica. Es de suponer que por inercia o falta de conocimiento más profundo una cantidad importante de usuarios busque cualquier cosa desde aquí, sea autor, título o materia.

En el primer listado, el de INDICES USADOS, cabe destacar que, hecha la salvedad de la búsqueda mixta por palabras clave, el índice con mayor frecuencia de uso es el de autores, lo cual resulta lógico si se piensa que los alumnos de los primeros cursos, más numerosos, suelen buscar fundamentalmente las bibliografías recomendadas por los profesores y cuentan con los datos de autor y título (también es muy alto el uso del índice de títulos). Les siguen el de materias y el de títulos de publicaciones periódicas que presuponen en general su utilización por usuarios más especializados .

Hay que señalar que las búsquedas que cuentan con resultados de más de 500 registros podrían considerarse como mal realizadas ya que deberían haberse refinado para ajustar su pertinencia . En la estadística de resultados de las búsquedas por palabras clave el porcentaje con estos resultados es el 17% mientras que en las búsquedas por materias es el 6% .

Las opiniones de los centros son, como cabría esperar, variadas, en general consideran satisfactoria la combinación de las posibilidades actuales de recuperación, aunque el grado de satisfacción varía y hay matices en función de las distintas situaciones, como es el caso de la Facultad de Informática que usa principalmente la búsqueda por palabra clave, puesto que su materia es muy cambiante, con innovaciones constantes y la terminología suele adoptarse directamente del inglés.

Los puntos que se suelen considerar más problemáticos para la búsqueda de los usuarios son el uso de los subencabezamientos, la adecuación de la terminología para que resulte más comprensible para cualquiera y la formación que requiere la

realización correcta de las búsquedas. En las conclusiones de este informe se desarrollan más todas estas cuestiones, así como diversas propuestas de actuación en relación con ellas.

EVALUACION DEL POSIBLE USO DE DESCRIPTORES

El uso de descriptores implica necesariamente el uso de tesauros puesto que un descriptor es por definición un término indizante de un tesoro y sólo puede entenderse dentro del contexto de éste, por ejemplo, el término: Independencia, en el Tesoro de términos relativos a relaciones raciales del CINDOC, se usa en lugar de autodeterminación dentro del marco de la política, mientras que en el Tesoro ISOC de Psicología, también del CINDOC, el mismo término se usa en lugar de autonomía personal. Teniendo en cuenta la amplitud temática de la documentación que utiliza y genera la UCM y la especialización de los campos a los que se refieren los tesauros, el número de éstos debería ser considerable, así como el de los índices correspondientes, uno por cada tesoro, ya que un mismo concepto puede tener significado y/o alcance diferente dependiendo del contexto y el enfoque de cada tesoro que lo utilice.

En la consulta realizada a los centros, nueve de ellos, es decir más del 64%, opinan que podrían usarse, pero sin abandonar el sistema actual en ningún caso, tres observan que sólo serían útiles para una minoría investigadora (profesores, alumnos de doctorado, etc,) y tampoco resultarían imprescindibles y a dos no les parece que interesen en absoluto en su área.

Cinco de los catorce afirman no conocer ningún tesoro útil para su materia, pero sólo entre los otros nueve mencionan diecisiete.

Millennium actualmente sólo dispondría de un posible índice libre y, en cualquier caso, aunque estuvieran todos libres, serían claramente insuficientes, por lo que la posibilidad de usar tesauros parece inviable en la práctica.

POSIBILIDAD DE USO DE DESCRIPTORES LIBRES

El uso de unos descriptores asignados libremente por el indizador, sin limitación de número ni control de vocabulario, es decir, de los llamados descriptores libres, está posibilitado por MARC21 a través de una etiqueta de materia, la 653 (Index term uncontrolled) definida para contener términos de indización que no corresponden a ningún sistema controlado

Esta etiqueta es repetible en el registro y , por lo tanto, permite asignar varios descriptores libres a cada documento.

La recuperación por este concepto implica la creación de un solo índice, del que posiblemente se podría disponer.

EVALUACION DEL USO DE DESCRIPTORES LIBRES

Existe mucha literatura acerca de las ventajas e inconvenientes que conlleva el uso del lenguaje libre y del lenguaje controlado. Los estudios suelen concluir que pueden ser complementarios, en general puede decirse que un lenguaje libre da mejores resultados cuanto más específica sea la búsqueda.

En cuanto a los costes o esfuerzo, existe una relación entre el input y el output de los sistemas de recuperación. En los sistemas de lenguaje controlado el coste y el esfuerzo se encuentran en la fase de entrada y en los de lenguaje libre en la de salida.

Las ventajas de la indización con descriptores libres radican en que permite representar conceptos más específicos que las palabras clave, disminuye la posibilidad de malas combinaciones y se pueden representar conceptos y temas aunque no aparezcan explícitamente en los documentos.

Los inconvenientes están en la falta de coherencia, duplicación de entradas al no existir control morfológico, errores ortográficos o tipográficos, el usuario se carga con la tarea de prever todos los términos que puede haber utilizado el indizador e incorporar específicos de un genérico si quiere que su búsqueda sea exhaustiva y las listas de resultados no pueden presentarse en ningún orden de pertinencia, puesto que los términos usados al indizar no tienen porqué estar presentes en ninguna parte del mismo documento y el sistema no tiene cómo valorarlos automáticamente.

La posibilidad de añadir descriptores libres a los registros bibliográficos de Millennium existe en cuanto que hay una etiqueta 653 específica para ello y requeriría un índice, con el que posiblemente se puede contar.

Debido a los problemas ya mencionados de duplicación de entradas y falta de coherencia, no parece lógico que el uso de descriptores libres sea sustitutivo del sistema actual de tesoro, sino, en todo caso podría ser complementario.

Si se decide usarlos, hay que tener en cuenta que el indizador debe conocer bien la materia para aplicar no sólo los términos correctos sino también todos sus posibles sinónimos y por otra, que el usuario también tiene que realizar búsquedas muy planificadas y exhaustivas si no quiere perder una parte considerable de la información.

Por otra parte habría que valorar qué ventajas aportaría, dado que supondría una carga de trabajo adicional. La que puede considerarse más rentable es la posibilidad de que la descripción sea tan específica como se quiera, mientras que las desventajas fundamentales son la imposibilidad de tener ningún tipo de control sobre las ambigüedades semánticas y los errores y crear confusión en los usuarios entre esta búsqueda y la de palabra clave.

La opinión de los Centros a este respecto es casi unánime, todos menos dos opinan que no deberían usarse fundamentalmente por dos motivos: constituirían una fuente de confusión tanto a nivel de falta de control como de forma de buscar de los usuarios y la carga de trabajo que supondrían no compensaría su utilidad. Uno de los dos Centros que afirman que podrían asumir esta carga de trabajo y que consideran muy interesante el uso de descriptores libres, también expresa sus reservas ante los problemas que acarrearía la falta de control.

CONCLUSIONES

Las conclusiones de este informe se pueden dividir en dos grupos:

Conclusiones referidas a la inclusión de nuevos sistemas de indización

- Existe un impedimento material básico para considerar la inclusión de descriptores controlados por tesauros o descriptores libres y es que cualquier opción en este sentido requiere el uso de múltiples índices y el programa no dispone en la actualidad más que de uno libre y, según la información proporcionada por Innovative al Servicio de Tecnología de la Información y Sistemas Bibliotecarios, parece que tampoco es posible aumentar el número.
- En el caso de los descriptores libres la opinión de los Centros es mayoritariamente negativa, tanto por considerar su aportación escasa y sus repercusiones perjudiciales para el catálogo y para los usuarios por el confusiónismo y la falta de referencias como por considerar desproporcionada la carga de trabajo que supondría en relación con su utilidad.
- La suma de la búsqueda por palabras clave y por los términos del Tesauro BUC, parece que cubriría suficientemente las necesidades de recuperación, sobre todo si se trabaja sobre los puntos que se indican en las siguientes conclusiones

Conclusiones referidas al sistema actual de materias

- Sería importante poder hacer registros de autoridad de subencabezamiento para que se facilite la claridad de las búsquedas, para ello sería importante insistir en la propuesta de que se habilite esta opción (contemplada en el MARC21) en GEUIN. Los representantes de la UCM ante este grupo ya tienen esta propuesta junto con otras presentadas por el Servicio de Proceso Técnico y Normalización.
- Se ha planteado como cuestión fundamental la formación de usuarios. Parece que el nivel de interés y satisfacción en la recuperación por materias es proporcionalmente mayor en los Centros en los que se consiguen buenos niveles de asistencia a los cursos de formación de usuarios y que cuentan con un servicio de asistencia a las búsquedas en la Biblioteca. Por lo tanto parece una cuestión importante el replanteamiento de la formación de usuarios tanto en lo referente a los contenidos de los cursos como al marketing de los mismos.
- Finalmente, otra cuestión que ha quedado planteada y a la que ya se está intentando buscar soluciones es la necesidad de usar en las interfaces públicas una terminología correcta pero comprensible para el usuario, que exprese la realidad ajustadamente y sin ambigüedad, pero que no suene a jerga profesional., así por ejemplo, se podría cambiar de materias a temas, de encabezamientos a términos principales y de subencabezamientos a

subdivisiones, así como dar explicaciones más comprensibles a las indicaciones del Tesauro y del Catálogo (algunas ya están hechas y otras están en vías de hacerse) y desarrollar las abreviaturas.

ANEXO 1

NORMAS ISO Y NORMAS UNE

NORMAS INTERNACIONALES Y NACIONALES

Así como para el campo de la descripción formal (Catalogación) existe una normalización muy abundante, el campo del análisis documental (Clasificación) está mucho menos normalizado y las diferentes opciones quedan prácticamente a criterio de las entidades o centros que van a realizarlo y dependiendo del tipo de estos centros y de las posibilidades de los sistemas informáticos de que dispongan. A continuación resumimos las normas ISO y UNE que afectan al contenido de este informe:

■ ISO-2788 1986 (E) Documentation - Guidelines for the establishment and development of monolingual thesauri. (Directrices para el establecimiento y el desarrollo de tesauros monolingües.)

Esta norma abarca algunos aspectos de la selección de términos de indización, los procedimientos para el control del vocabulario y, específicamente, el modo de establecer relaciones entre estos términos (particularmente aquellas relaciones que *a priori* se utilizan en los tesauros), así como la inclusión y supresión de términos, los métodos de compilación, la forma y el contenido de los tesauros, el uso de la automatización en el procesamiento de los datos, etc.

■ ISO 5963 1985 (E) Documentation - Methods for examining documents, determining their subjects, and selecting indexing terms. (Métodos para examinar los documentos, determinar su materia y seleccionar los términos de indización.)

Se describen los métodos para examinar y analizar los documentos, determinar la materia que tratan y seleccionar los términos de indización apropiados. Esta norma se limita a las etapas preliminares de la indización y no incluye las prácticas de un tipo específico de sistema de indización, ya sea precoordinado o postcoordinado. Las técnicas descritas en esta norma pueden ser usadas por cualquier institución de información donde el proceso de indización se realice de forma manual; por el contrario, no se aplican en aquellas instituciones que empleen técnicas de indización automatizada. Esta norma tiene como propósito guiar a los indizadores durante las etapas de análisis de documentos e identificación de conceptos esenciales en la descripción de la materia.

■ ISO 5964 1985 (E) Documentation for the establishment and development of multilingual thesauri. (Guía para el establecimiento y desarrollo de tesauros multilingües.)

Las reglas ofrecidas en esta norma deben utilizarse en conjunto con la norma ISO 2788 (Guía para el establecimiento y desarrollo de tesauros monolingües), pues la

mayoría de los métodos y recomendaciones consideradas en ésta son igualmente válidas para los tesauros multilingües. La presente norma se considera como un paso fundamental en el perfeccionamiento de la recuperación de la información y el logro de la compatibilidad entre los tesauros producidos por instituciones que indizan con términos seleccionados a partir de más de dos lenguajes naturales (idiomas). Su contenido abarca los problemas multilingües que pueden surgir durante la creación de un tesoro convencional.

■ **UNE 50113-5. Documentación e información. Vocabulario. Adquisición, identificación y análisis de documentos (Equivale a la norma ISO 5127/3A-1981)**

Aporta las siguientes definiciones:

- Análisis de contenido: análisis documental que tiene por objeto facilitar los datos que caracterizan el contenido de un documento de una forma clara y concisa.
- Indización: operación destinada a representar los resultados del análisis de contenido de un documento o de una parte del mismo mediante elementos de un lenguaje documental o natural para facilitar la recuperación.
- Asignación de términos de indización: elección y atribución de términos, aparezcan o no éstos en el texto, para representar documentos o datos de acuerdo con ciertas reglas.
- Extracción de términos: proceso por el cual los términos de indización se toman del texto que se indiza.

■ **UNE 50113-6, Febrero 1998 (Equivale a la Norma ISO 2382) Documentación e información. Vocabulario. Lenguajes documentales.**

Precisa la terminología, estableciendo definiciones que aseguren la equivalencia correcta en todos los idiomas y proporcionando un significado único a todos los lectores. Para ello se basa en la utilización habitual de los términos en vocabularios nacionales relevantes, en documentos de entidades internacionales (ISO, IFLA, FID, UNESCO, etc.) y en normas nacionales:

- Lenguaje documental: lenguaje formalizado, usado para caracterizar los datos o el contenido de los documentos y permitir su almacenamiento y recuperación.
- Sistema de clasificación: lenguaje documental destinado a la representación estructurada del documento o datos mediante símbolos de clase y de los términos correspondientes para permitir el enfoque sistemático.
- Tesoro: vocabulario controlado de términos que tienen entre sí relaciones semánticas y que cubre uno o más campos especializados del conocimiento.
- Elementos básicos de los lenguajes documentales:
 - Palabra clave: palabra o grupo de palabras seleccionadas del título o del texto de un documento que caracterizan su contenido y permiten su recuperación.

- Descriptor: término de un tesoro que puede emplearse para representar un concepto contenido en un documento o en una petición de recuperación.
- No descriptor: término de un tesoro que no debe emplearse para representar un concepto pero que remite a uno o varios descriptores que pueden utilizarse en su lugar.
- Identificador: nombre utilizado como descriptor.
- Calificador: símbolo o palabra utilizado para diferenciar los distintos significados de los homógrafos.

■ **UNE 50-121-91 (Equivale a la Norma ISO 5963 de 1985) Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización.**

- Materia: cualquier concepto o combinación de conceptos que representa el tema de un documento
- Término de indización: la representación de un concepto en forma de un término derivado del lenguaje natural, preferentemente un sustantivo simple o compuesto o de un código de clasificación.
- Indización: extracción de los conceptos de un documento mediante un proceso de análisis intelectual, transformándolos en términos de indización y valiéndose de herramientas de indización: tesauros, sistemas de clasificación, etc. Tiene tres etapas:
 - Examen del documento y determinación de su contenido.
 - Identificación y selección de los conceptos principales del contenido.
 - Selección de los términos de indización.

■ **Guidelines for Subject Authority and Reference Entries.** Working group on "Guidelines for Subject Authority Files" of the Section on Classification and Indexing of the IFLA Division of Bibliographical Control. München, etc. : K.G.Saur, 1993. Solamente contemplan el caso de los encabezamientos de materia.

ANEXO 2

INFORMACIÓN DE GESTIÓN SOBRE BÚSQUEDAS EN LOS CATÁLOGOS PÚBLICOS DE LA UNIVERSIDAD COMPLUTENSE

Esta información ha sido tomada de *Millennium Web Management Reports. User Access. Patron Searches*. Algunos datos que comentaremos a continuación de la tabla correspondiente, resultan dudosos y hay una consulta abierta con Innovative a la que estamos esperando respuesta.

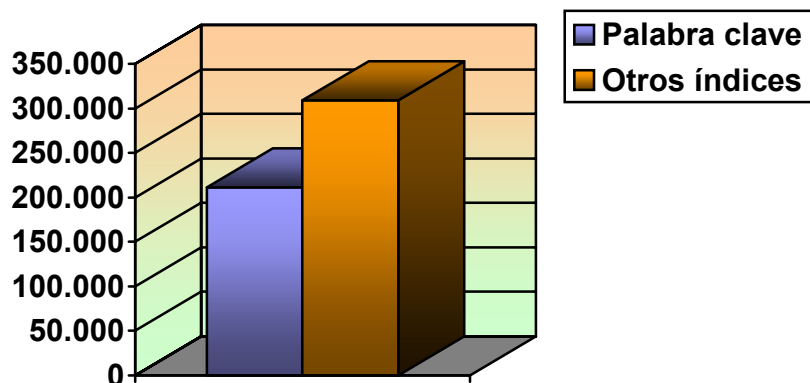
INFORME DE GESTIÓN SOBRE BÚSQUEDAS EN CATÁLOGOS PÚBLICOS

Informe de actividad de búsquedas	
Desde 4-3-06 a 3-05-06	
Número de búsquedas del usuario	550.623
Número de búsquedas sugeridas por el sistema *	142.477
Número de registros recuperados	450.157.846
Número de búsquedas limitadas	16.864
Número de registros exportados	17.446
Número de pantallas solicitadas	1.165.080

* Las búsquedas sugeridas por el sistemas son las que se producen cuando en una búsqueda por título o por materia no hay resultados. En ese caso el sistema responde ofreciendo la posibilidad de buscar por palabra clave y, si se acepta, mostrando un formulario para esta búsqueda con todas sus opciones. Estas búsquedas incrementan, como es lógico, la estadística de búsquedas por palabra clave.

Si en una búsqueda por autor en la que se han utilizado dos elementos (nombre y apellido o dos apellidos, o incluso dos nombres) tampoco hay resultados, el sistema ofrece la posibilidad de buscar permutando el orden.

Índices usados y resultados de las búsquedas		
Índices	Número total	Porcentaje
Número de registro	1340	0,24%
Palabras clave	211.420	38,40%
Autores	125.267	22,75%
Signatura variante CDU	1.194	0,22%
Materias	52.148	9,47%
Colecciones	2.910	0,53%
Signatura libre	97	0,02%
Números clasificación	2.885	0,52%
Signatura currens	2.252	0,41%
ISN	16.349	2,97%
Nombre del recurso	1	0,00%
Signatura literal	25	0,00%
Número de control	2	0,00%
Profesores/tutores	2.581	0,47%
Asignaturas	4.674	0,85%
Título de publicación periódica	21.759	3,95%
Títulos	105.109	19,09%
Lugar de impresión (Fondo Antiguo)	378	0,07%
Descriptorios	227	0,04%
TOTAL	550.618	100%



Comparación de tipos de búsquedas

El concepto "Otros índices" incluye la suma de las búsquedas realizadas por los índices en cuyo contenido también se realiza la búsqueda por palabra clave: Autores, Materias, Colecciones, Título de publicaciones periódicas y Títulos. La comparación de las búsquedas por palabra clave con la de cualquier otro índice por separado, no tendría sentido.

RESULTADOS DE BUSQUEDAS DEFINIDAS POR EL USUARIO

Índices usados y resultados de las búsquedas - Por palabra clave -		
	Nº total	Porcentaje
Búsquedas que recuperan 1 registro	101.175	47,85%
Búsquedas que recuperan de 2-8 registros	28.525	13,49%
Búsquedas que recuperan de 9-30 registros	18.935	8,96%
Búsquedas que recuperan de 31-99 registros	13.997	6,62%
Búsquedas que recuperan de 100-499 registros	14.457	6,84%
Búsquedas que recuperan de 500-4999 registros	16.102	7,62%
Búsquedas que recuperan 5000 o más registros	18.227	8,62%
Total de búsquedas que recuperaron registros	211.418	100%
Total de registros recuperados y promedio por búsqueda	422.466.007	1.998
Búsquedas sin recuperación directa	2	0,00%
Total de Búsquedas	211.420	100%

Esta estadística tiene los siguientes puntos en los que hay que insistir:

- En el total se incluyen, además de las búsquedas realizadas directamente por el usuario por palabra clave, las sugeridas por el sistema al no encontrar resultados buscando por otros criterios, como se ha señalado en el informe de actividad de búsquedas, y las de autor-título en la parte correspondiente al título (busca por palabra clave en el campo de título).
- El epígrafe " Total de búsquedas que recuperaron registros" se muestra como el 100%, cuando está comprobado que realmente hay búsquedas sin resultado.
- Las búsquedas que recuperan 1 registro figuran como el 47,85% y podemos afirmar que esto es imposible ya que, dadas las características de este tipo de búsqueda puede darse que no haya resultados cuando se busca una errata o un dato muy concreto y raro, pero no hemos podido detectar búsquedas con un solo resultado. Sobre la interpretación de estos datos dudosos se ha abierto una consulta a Innovative.
- Finalmente hay que señalar que el número de consultas que ofrecen resultados con más de 500 registros y que por lo tanto son poco específicas, supone el 16,24%

Índices usados y resultados de las búsquedas - Por materias -		
	Nº total	Porcentaje
Búsquedas que recuperan 1 registro	25.423	28,75%
Búsquedas que recuperan de 2-8 registros	2.798	5,37%
Búsquedas que recuperan de 9-30 registros	1.941	3,72%
Búsquedas que recuperan de 31-99 registros	2.030	3,89%
Búsquedas que recuperan de 100-499 registros	2.759	5,29%
Búsquedas que recuperan de 500-4999 registros	2.620	5,02%
Búsquedas que recuperan 5000 o más registros	584	1,12%
Total de búsquedas que recuperaron registros	38.155	73,17%
Total de registros recuperados y promedio por búsqueda	11.353.625	297
Búsquedas sin recuperación directa	13.993	26,83%
Total de Búsquedas	52.148	100%