

## GUÍAS PARA LA SELECCIÓN Y UTILIZACIÓN DE TÉRMINOS DE BÚSQUEDA EN BASES DE DATOS CON CAMPOS DE LENGUA NATURAL

PEDRO FALCATO

---

**Resumen:** Los campos que contienen lengua natural son de gran valor como puntos de acceso a los contenidos de las bases de datos. Sin embargo, los usuarios de los servicios de información se encuentran con diversos problemas cuando intentan utilizarlos. Se elaboró un esquema descriptivo de las acciones que pueden realizarse durante un proceso de búsqueda de este tipo; se registraron también dificultades, desviaciones y carencias del proceso de selección y combinación de términos tal como lo efectúan los usuarios finales. Finalmente se formularon recomendaciones para la elaboración de guías de ayuda.

**Palabras clave:** Lengua natural; Selección de términos de búsqueda; Combinación de términos de búsqueda; Ayudas; Usuarios.

**Abstract:** Fields containing natural language terms are valuable as access points to databases contents. However, information services users have varied problems when trying to make use of them. A description scheme was developed considering the possible actions during a search process of this kind. Difficulties, deviations and faults of the terms selection and combination processes were recorded. Finally, recommendations were carried out in order to produce search aids.

**Keywords:** Natural language; Search terms selection; Search terms combination; Helps; Users.

---

- Trabajo realizado para el proyecto: Evaluación de fuentes de información en relación con la búsqueda de información y la intervención del referencista, en el área de tecnología industrial (en ejecución en CIME-CID).
- Universidad de Buenos Aires, Facultad de Filosofía y Letras, Instituto de Investigaciones Bibliotecológicas. Departamento de Bibliotecología y Ciencia de la Información.
- Instituto Nacional de Tecnología Industrial, CIME-CID.

Correo electrónico: pfalcato@abaconet.com.ar

Artículo recibido: 04-10-99. Aceptado: 10-05-00.

*INFORMACIÓN, CULTURA Y SOCIEDAD*. No. 3 (2000) p. 33-46

©Universidad de Buenos Aires. Facultad de Filosofía y Letras. Instituto de Investigaciones Bibliotecológicas (INIBI), ISSN: 1514-8327.

## Introducción

La difusión del acceso a bases de datos por parte de los usuarios finales, tanto de manera local como remota, ha generado importantes novedades en los modos de acceso a la información.

Desarrollar teorías y técnicas que permitan gestionar información y conocimiento en ambientes saturados de información es, indudablemente, un desafío característico de nuestra época.

En este contexto, los campos que contienen lengua natural son de gran valor como puntos de acceso a los contenidos de las bases de datos. No es novedoso que la búsqueda en títulos, resúmenes, etc. es conveniente en muchos casos ya que, por ejemplo, posibilita ampliar la recuperación cuando se desconocen las características del vocabulario controlado, un aspecto temático no está representado por éste de una manera adecuada a las necesidades del caso, etc. A las bases bibliográficas se han agregado las de texto completo y, en los últimos tiempos, una gran variedad de recursos en Internet, detectables mediante motores de búsqueda que trabajan, fundamentalmente, operando sobre la lengua natural presente en las páginas. Otro punto importante es que la mayoría de las tácticas adaptadas a estas condiciones puede emplearse de modo iterativo dentro de una estrategia que abarque diversas bases o documentos.

Para lograr una recuperación satisfactoria en estas condiciones, la selección y combinación de términos de búsqueda son etapas de fundamental importancia. Sin embargo, muchos usuarios desconocen cómo aprovechar las posibilidades que se les presentan. Ciertas técnicas que lo permiten pueden resultar dificultosas para ellos. En cambio otras están al alcance más o menos inmediato del público en general. En todo caso, el diseño de las ayudas juega un importante papel. Estas pueden estar presentes como una opción de la interfaz, o también como archivos que contienen listas de consejos (*search tips*) o de dudas frecuentes (*frequently asked questions* o *FAQs*), etc.

Por otro lado, desde el punto de vista de los países que tienen su infraestructura de información en un nivel de desarrollo intermedio o bajo, es común tener que enfrentarse con un reto muy distinto: la gestión en un ambiente informativamente pobre. No se dispone de todas las herramientas bibliográficas necesarias o es costoso acceder a ellas, etc. Y cuando se accede, con frecuencia los resultados se refieren a documentos que están fuera del alcance inmediato. También quedan excluidas publicaciones locales sin difusión internacional. Se hace necesario entonces aprovechar al máximo los recursos disponibles a nivel local, para lo cual conviene desarrollar la catalogación analítica. En ella la indización basada en la lengua natural presente en títulos, resúmenes o notas de contenido, es muy útil para la recuperación y resulta complementaria de la que se realiza con vocabulario controlado.

Inclusive, en no pocas circunstancias, la escasez de personal y de recursos económicos impiden realizar un análisis temático profundo de los documentos. Esto es particularmente cierto cuando la complejidad y características multidisciplinarias del área cubierta se asocian con la imposibilidad de contar con un equipo de indizadores especializados en cada uno de los temas. En estos casos, el empleo de la lengua natural permite generar herramientas útiles aun cuando no se puede ofrecer un servicio completo que incluya la utilización de un vocabulario controlado. Esta opción, si bien más limitada, es aplicable en tales situaciones porque no deben afrontarse los gastos relacionados con la asignación de términos, construcción de tesauros, etc. (Oard, 1997)

En el Centro de Investigación y Desarrollo de Métodos y Técnicas para Empresas Industriales, División CIME Información y Documentación (CIME-CID) se compila la base TEC, que contiene registros correspondientes a artículos seleccionados en publicaciones periódicas existentes en la Biblioteca. La indización de la misma se basa en los títulos de dichos documentos, traducidos al español si están redactados en otras lenguas y enriquecidos con expresiones aclaratorias cuando no representan adecuadamente el contenido. Se agregan en otro campo categorías temáticas generales, para contextualizar los términos. La cobertura es multidisciplinaria y no se dispone de un equipo de clasificadores especializados por áreas.

Tanto por medio de observaciones espontáneas como de consultas, mensajes y comentarios recibidos, se fueron detectando algunos problemas característicos que deben enfrentar quienes realizan búsquedas en TEC. Casos análogos se encontraron al orientar a usuarios de otras bases de datos con campos en lengua natural y de motores en Internet.

A partir de lo antedicho se decidió desarrollar el presente trabajo, que tiene por objeto proponer orientaciones destinadas a quienes redactan guías para la selección y utilización de términos de búsqueda. Estas, además de integrarse a las interfaces, podrían utilizarse como herramientas en cursos que se dictan a estudiantes de bibliotecología o en el entrenamiento de usuarios.

## **Metodología**

Para hallar soluciones referidas a esta problemática, se reunieron elementos para conformar un modelo descriptivo de las acciones que pueden realizarse durante un proceso de búsqueda en lengua natural. Se consideró adecuado adoptar una aproximación cualitativa; se reunió y categorizó información sobre tipos y posibles causas de las dificultades para obtener una recuperación adecuada. No se optó por el análisis cuantitativo de estos datos, considerando que la oferta de elementos de ayuda se regiría por la existencia y la naturaleza de las operaciones y problemas más que por su número.

En primera instancia se revisaron antecedentes que corresponden a ayudas ya implementadas (AltaVista, 1999; Carnegie Mellon University. Lycos, 1999; Northern Light Technology, 1999; Telecom Soluciones, 1999; Wired Digital, 1999; Yahoo, 1999).

Fueron hechas además 20 entrevistas semiestructuradas, tanto a usuarios finales como a bibliotecarios y otros profesionales dedicados a la realización de búsquedas. El material reunido en las entrevistas a las dos últimas categorías, por su mayor claridad conceptual, se utilizó junto con los antecedentes bibliográficos para generar un primer esquema de operaciones usuales. Este fue enriquecido y afinado considerando los datos correspondientes a usuarios finales y las observaciones que se mencionan más adelante.

Se plantearon potenciales necesidades de información para que los entrevistados esbozaran estrategias de búsqueda, explicando por qué tomaban las decisiones expresadas. También se indagó acerca del conocimiento general que cada uno tenía acerca de los operadores usuales que pueden emplearse para las búsquedas en campos de lengua natural, así como de las formas de sintaxis más frecuentes.

En el caso de los usuarios finales no se plantearon los temas utilizando un vocabulario técnico, sino que se presentaron de manera sencilla las consignas, y se utilizaron ejemplos y analogías apropiadas para aclarar sus alcances.

Los problemas que se propusieron se seleccionaron para posibilitar la detección de elementos claves, tales como la utilización iterativa de tácticas y el refinamiento de búsquedas a partir del examen de resultados parciales, la supresión de palabras poco significativas (dentro de un determinado contexto) al redactar las expresiones de búsqueda, la tendencia a utilizar un vocabulario rico o limitado, el manejo de relaciones género-especie, etc.

Asimismo se analizaron 50 observaciones de situaciones de búsqueda en CIME-CID, estudiando los objetivos registrados en los formularios de consulta que se emplean en el servicio de Referencia, las acciones llevadas a cabo, los comentarios expresados por los bibliotecarios o por los usuarios, las características de las fuentes de información utilizadas y finalmente los resultados.

## **Resultados**

El esquema obtenido se expone a continuación. Es necesario tener en cuenta que no todas las acciones enunciadas pueden llevarse a cabo en cualquier sistema, ya que cada uno ofrece distintas posibilidades.

### *Operadores usuales*

Se revisó documentación correspondiente a diversos sistemas. (AltaVista, 1999; Carnegie Mellon University. Lycos, 1999; Deharbe, 1999; Dialog

### Ejemplos de símbolos utilizados para representar operadores

Operador	Acción general	Ejemplos de simbología utilizada
Intersección de conjuntos de términos	Selecciona documentos que contienen ambos términos que se encuentran a los lados del operador .	and & *
Suma de conjuntos de términos	Selecciona documentos que contienen al menos uno de los términos que se encuentran a los lados del operador.	or   +
Diferencia de conjuntos de términos	Excluye los documentos que contienen el término que se encuentra a continuación del operador.	not and not ! ^
Proximidad de términos	Selecciona documentos donde los términos que se encuentran a ambos lados del operador se encuentran cercanos uno del otro (adyacentes, con hasta un determinado número de palabras en el medio, etc.)	near ~ adj \$ . (f) (g) [con ligeras variantes en su modo de acción]
Frase exacta	Selecciona documentos en los que está presente la expresión compuesta, en el orden indicado.	“ ”
Paréntesis	Indica niveles de precedencia en las operaciones.	( )
Inclusión - exclusión	Indica que el término que sigue al operador debe estar presente (+) en los documentos seleccionados, o ausente (-) en los mismos.	+ -
Comodín	Se intercala en un término de búsqueda para indicar que en esa posición puede haber cualquier carácter (o cadena de caracteres, según el caso)	? * % [con ligeras variantes en su modo de acción]
Truncamiento	Se ubica a continuación de un fragmento de una palabra, para indicar que se buscan documentos que contengan términos que comiencen con la raíz expuesta, sin importar cual sea la terminación.	? * \$

Corporation, 1998; Martínez et. al., 1999; Northern Light Technology, 1999; Telecom Soluciones, 1999; Wired Digital, 1999; Yahoo, 1999).

Los operadores de inclusión pueden usarse para producir un efecto semejante a la intersección de conjuntos. Los de exclusión, a la diferencia de conjuntos.

Es posible lograr resultados bastante parecidos con los operadores de proximidad, por un lado, y las frases exactas por otro; ambos ubican los términos aislados en un contexto y pueden verse como formas más exigentes de la intersección de conjuntos, ya que se requiere que las palabras no sólo estén presentes sino además en determinada posición relativa.

Los comodines y el truncamiento sustituyen la enumeración de términos emparentados ortográficamente.

En algunos sistemas se admite la presentación de consultas redactadas como preguntas en lengua natural. Sin embargo, por el momento los resultados no siempre son satisfactorios.

### *Acciones usuales*

Búsquedas de algunos documentos o fragmentos acerca de un tema (no exhaustivas)

- Seleccionar alguna palabra importante que represente, aunque sea de manera parcial, el tema; en lo posible debe ser poco frecuente fuera de ese contexto.
- Si el sistema lo admite, puede intentarse:
  - Ingresar varias palabras sin indicar operadores específicos.
  - Utilizar preguntas en lengua natural.

Realimentación de la búsqueda (si no se halló algún registro, documento o fragmento que corresponda a lo esperado):

- Intentar con otra palabra (sinónimo, o término de significado más general o más específico, según las necesidades)
- Revisar los resultados para detectar nuevas pistas, términos a emplear, etc.
- Revisar la ortografía de las palabras empleadas.
- Si luego de varios ensayos no se encuentra información, intentar con algunas de las acciones empleadas para búsquedas exhaustivas.

Búsquedas temáticas que intentan ser exhaustivas

- Determinar expresiones que describan el tema.
- Analizar las palabras presentes y sus relaciones de proximidad mutuas.
- Buscar sinónimos o cuasi-sinónimos (recordando incluir las variantes ortográficas), así como términos relacionados (aunque no se utilicen para nombrar al tema en sí; por ejemplo, pueden elegirse nombres que correspondan a aspectos más específicos del mismo o asuntos muy frecuentemente relacionados). Repetir los dos primeros pasos.
- Seleccionar un conjunto mínimo de términos que representen todas las variantes consideradas. No es necesario incluir todas las palabras que conforman cada enunciación: se prefieren aquellas que puedan estar menos representadas en registros que correspondan a temas diferentes, o sea menos ambiguas.

- Descartar las palabras que aportan poco sentido en el contexto (por ejemplo, la palabra Instituto en una base de datos sobre Institutos).
- Seleccionar raíces que representan varios términos.
- Descartar (si se estima que disminuirán demasiado la precisión) las raíces que pueden referir a expresiones no relacionadas con el tema. Si es necesario, se las reemplaza por una enumeración de términos completos.
- Descartar raíces demasiado breves, ya que pueden provocar demoras y otros inconvenientes como la inclusión de términos no deseados.
- Conectar las palabras que adquieren un sentido particular estando cerca de otras, por medio de operadores de proximidad o bien integrarlas en una frase.
- Efectuar una suma de conjuntos con cada grupo de sinónimos, cuasi-sinónimos o términos relacionados, conformando clases.
- Asociar las diversas clases que deben estar presentes mediante intersección de conjuntos, para restringir el resultado.
- Si hay relaciones de precedencia, utilizar los paréntesis que correspondan.
- Si es posible, reducir la longitud de las expresiones usando paréntesis, mediante la propiedad distributiva de los operadores.
- Si es necesario, restringir el resultado mediante resta de conjuntos; esta opción debe evitarse cuanto sea posible, ya que puede llevar a la pérdida de información relevante (quizá los términos restados aparezcan en contextos que no justifican excluir los registros que los contienen; por ejemplo enumeraciones cualesquiera, comparaciones o incluso frases donde se aclara que los conceptos representados por ellos están fuera del enfoque del documento).
- Si se desea restringir mediante intersección de conjuntos, pero no hay certeza de que los términos que conforman esa clase estén siempre presentes cuando se trata el tema, es conveniente no hacerlo salvo que el resultado sea demasiado extenso. En general es preferible revisar algunos registros más y no perder información relevante.

#### Realimentación de la búsqueda:

- Estudiar si hay palabras que produjeron problemas (demasiado generales, demasiado específicas, polisémicas, etc.)
- Si hay frases, revisar el espaciado, la puntuación, etc.
- Revisar la ortografía
- Revisar los resultados para detectar nuevas pistas, términos a emplear, etc.

#### Si hay exceso de recuperación :

- Estudiar si es factible reemplazar términos de alcance general por otros más específicos, o bien suprimirlos.

- Estudiar si se puede acotar el resultado agregando relaciones de proximidad u otras restricciones.

Si hay escasa recuperación:

- Agregar sinónimos, términos más específicos o relacionados.
- Suprimir restricciones (operadores de proximidad, intersecciones o diferencias de conjuntos)
- Buscar palabras que representen temas más generales.

Búsquedas de documentos o fragmentos específicos, de los cuales se conoce parte del texto

- Emplear una expresión que figure en el texto. Puede ser una sola palabra poco frecuente, sobre todo si la base de datos es pequeña; también una expresión compuesta mediante la intersección de conjuntos, operadores de proximidad o una frase exacta.

## Discusión

El paradigma bibliográfico se basa en la certeza y el orden. Los problemas de los usuarios, en cambio, se caracterizan por la incertidumbre y la confusión (Kuhlthau, 1991). El autor de un texto expresa sus ideas mediante palabras que, volcadas a una base de datos o a un documento electrónico, pueden ser empleadas para la recuperación. Esa expresión tiene características muy variadas en cuanto a su estilo, vocabulario, e incluso a nivel ortográfico (Brooks, 1998). El usuario, por su parte, formula una expresión de consulta con términos que el sistema intenta aparear con los del texto. Sin embargo, de hecho, la relación es simétrica sólo en la superficie. La experiencia del usuario es fenomenológicamente distinta ya que, salvo que esté buscando un ítem previamente identificado, intenta describir algo que no conoce; describe una grieta en su conocimiento y sólo puede adivinar cómo debería ser el relleno de esa grieta. O bien apunta a un área más amplia que su tema de interés específico, para intentar seleccionar luego lo que le parezca relevante dentro del conjunto obtenido (Bates, 1996). Para ello puede decidirse por una estrategia heurística, altamente interactiva, efectuando diversos intentos y revisando los resultados parciales (Marchionini, 1989), que pueden ser muy extensos. La lectura de hojear, fundamental en una estrategia de estas características, minimiza la carga cognitiva, trasladándola al sistema perceptual. Esta característica es atrayente para muchas personas porque los seres humanos normalmente buscan el camino de menor carga cognitiva (Yuan, 1997).

Por su parte, los bibliotecarios que se dedican a la búsqueda de información con frecuencia deben tender a lograr una recuperación exhaustiva;



con esto pueden ofrecerle al usuario un conjunto completo de ítems para que seleccione los que desee y además tenga un panorama de todo lo que está disponible.

Los expertos utilizan de manera complementaria la lengua natural y los vocabularios controlados, procurando alcanzar un nivel óptimo de recuperación. Muchos usuarios finales, por su parte, ni conocen las características de estas herramientas ni están predispuestos a estudiarlas. En tales situaciones las posibilidades que presentan los textos redactados en lengua natural pueden ser convenientes (Oard, 1997), ya que la variedad del léxico permite recuperar información con estrategias de búsqueda muy simples.

Por otro lado, resulta evidente que es casi imposible obtener de esta manera una proporción elevada de los ítems potencialmente útiles, mientras no se desarrollen medios más poderosos. En efecto, los usuarios suelen encontrarse con dificultades cuando intentan formular expresiones de búsqueda completas (Marchionini, 1989). Pero muchas veces sólo necesitan conseguir unos pocos documentos o fragmentos sobre un tema y se detienen en cuanto los encuentran. Esto quizás se relacione con los resultados de estudios como el realizado recientemente en la National Library of Medicine (Estados Unidos) para evaluar el impacto de MEDLINE, que indican escasa realimentación durante el desarrollo de las búsquedas para reajustar las estrategias (National Library of Medicine, 1997).

La aún escasa normalización de las interfaces y de los lenguajes de búsqueda produce también gran confusión entre los usuarios finales. Es muy común observarlos intentando utilizar en un sistema estrategias o tácticas adaptadas a las características de otro.

Las técnicas más empleadas por los usuarios son sencillas. Hay pocas personas que restringen adecuadamente empleando la intersección de conjuntos. Son todavía menos las que emplean sinónimos y términos relacionados que, reunidos, permitirían representar los conceptos de manera más apropiada e incrementar la recuperación (Sewell y Teitelbaum, 1986; Yuan, 1997).

Según Iivonen y Sonnenwald (1998), se ha utilizado generalmente para la elección de términos de búsqueda un modelo de analogía con la traducción, el cual no incentiva la búsqueda de nuevas palabras, sino el reemplazo de un vocablo por otro. Proponen uno que consideran mejor adaptado al comportamiento de los especialistas en búsquedas: la reunión de expresiones diversas mediante la navegación a través de distintos discursos, entre ellos los contenidos de documentos, las prácticas de indización, las solicitudes de búsquedas de los clientes, los contenidos y estructuras de las bases de datos y las experiencias obtenidas en búsquedas previas. Como sea, en muchos casos es decisiva, además de la adecuación temática, la variedad de los términos que se emplean. Hay que considerar también las variantes morfológicas, por ejemplo: singulares, plurales, masculinos y femeninos (Bates, 1996).

Otra concepción que puede resultar problemática es la que establece una analogía entre la interacción con un sistema automático, por un lado, y una conversación por el otro. Esta apela al modelo mental que las personas tienen acerca del diálogo entre seres humanos, en el cual no es necesario enumerar todos los términos relacionados con un determinado asunto, las diferencias morfológicas y sintácticas se dan por sobreentendidas, etc. Los participantes de la conversación, sin embargo, entienden de qué se está hablando. Los sistemas de computación, en cambio, suelen manejarse con reglas mucho menos flexibles (Bates, 1996).

En las estrategias planteadas por los usuarios finales, las operaciones de truncamiento son más infrecuentes que las booleanas, y más aún las orientadas a la proximidad de términos.

Por otro lado, es común que se utilicen frases (sin añadir operadores de ningún tipo) en sustitución de palabras integradas en ecuaciones de búsqueda, aunque se trate de sistemas que no están preparados para aceptarlas. Esto refleja un modelo mental muy pobremente definido acerca de las características de los sistemas que se emplean (Marchionini, 1989); también indica una preferencia por generar nombres para identificar textos combinando varias palabras en una expresión, así como por manifestar necesidades poniendo más énfasis en el planteo de conceptos que en la riqueza del vocabulario; los términos de búsqueda aislados se consideran representaciones poco confiables de esos conceptos (Collantes, 1995). Con respecto a las variantes morfológicas, ortográficas, etc., los usuarios tienden a suponer que la forma de una palabra que ellos introducen coincidirá exactamente con las ocurrencias en la base de datos (National Library of Medicine, 1997)

Las técnicas sencillas empleadas por la mayoría de los usuarios se vuelven inapropiadas cuando se trabaja con bases de datos de gran magnitud, hoy cada vez más frecuentes. Los enormes conjuntos de resultados que pueden obtenerse mediante estrategias muy simples en un motor de búsqueda en Internet sirven como ejemplo (Zorn et al., 1996). Más aún, muchas veces los términos seleccionados aparecen en contextos totalmente diferentes al esperado.

De lo expuesto hasta aquí se desprende la conveniencia de que algunas reglas provenientes del metaconocimiento de los especialistas en búsquedas (Connell, 1995) sean incorporadas a una lista de consejos u otra forma de ayuda accesible desde la interfaz.

La mayor parte de los usuarios finales de los sistemas de información tiende a aprender sólo lo necesario para efectuar búsquedas simples de manera razonablemente rápida. Supone que continuar instruyéndose y desarrollando sus habilidades es innecesariamente trabajoso. Por otro lado, la gran cantidad de variantes en las reglas de sintaxis, modos de operación, símbolos utilizados, etc., así como la elevada frecuencia de los cambios y actualizaciones, producen confusión y fortalecen la percepción general de que no conviene emplear demasiado tiempo en profundizar acerca de los mismos.

Con algunas pautas generales sobre el manejo del sistema, los usuarios pueden realizar búsquedas con cierto nivel de satisfacción desde el comienzo. De hecho, algunas posibilidades de los sistemas sólo en raras ocasiones son aprovechadas, salvo por los expertos. Una manera de difundir su existencia para, que se las utilice cuando sea necesario, es presentar los componentes básicos del sistema en un planteo inicial e introducir el resto posteriormente, cuando los usuarios ya pudieron trabajar con búsquedas reales (Yuan, 1997).

Tomando como ejemplo los motores de búsqueda en Internet, éstos en general dividen sus ayudas en una sección para principiantes y otra para expertos. Es común que cada una de esas secciones tenga relación con una interfaz distinta. La primera presenta menos variedad de posibilidades en lo que concierne a comandos y operadores, siendo su uso relativamente sencillo; suele haber otra con características más complejas. Sin embargo no siempre la interfaz avanzada es considerada preferible; de hecho, la documentación de AltaVista recomienda no utilizarla salvo en casos especiales, que no identifica (AltaVista, 1999).

Si el sistema permite ingresar consultas en forma de frases o preguntas en lengua natural, o bien introducir una serie de palabras sin especificar conexiones entre ellas, las secciones de ayuda básica comienzan por plantear estas posibilidades como la primera opción. La presentación del resto de la ayuda en algunos casos se basa en una breve presentación de los operadores y sus características. Al comienzo (o en los primeros niveles de acceso de las ayudas hipertextuales) se sitúa la explicación de los temas más importantes y generales, en especial los operadores básicos. Otro enfoque se basa en la oferta de listas de consejos, que intentan responder a dudas y dificultades usuales sobre la sintaxis de las expresiones de búsqueda y la selección de términos. Hay diversas variantes (*FAQs*, *search tips*, etc.). Las listas de problemas usuales o *FAQs* son convenientes porque los usuarios, en su mayoría, no están interesados en las generalidades y aspectos teóricos de los lenguajes de recuperación, sino en obtener mejores resultados en casos concretos. Es práctico que pueda accederse a los ítems de estas listas desde una tabla de contenido de modo hipertextual, y que también puedan ser leídos secuencialmente. En general, se debe avanzar hacia el diseño de interfaces que reduzcan la necesidad de confiar en la memoria y permitan aprender nuevos comandos y posibilidades a medida que se los necesite (Yuan, 1997). Las alternativas expuestas no son mutuamente excluyentes.

El ofrecimiento de acceso a explicaciones que pueden servir como guía, sin embargo, no es una solución para todos los casos; muchas personas casi no leen las ayudas, ni siquiera las más simples y breves, sino que prueban los sistemas introduciendo términos y expresiones directamente, sin detenerse en analizar si existe una manera de obtener mejores resultados. Por eso en los sistemas estudiados es una práctica común poner algún ejemplo breve y representativo

en la interfaz, muy próximo al lugar donde deben escribirse los términos de búsqueda.

Si el sistema contiene poca información, no será necesario presentar demasiadas técnicas para la restricción de resultados, ya que difícilmente serán empleadas.

En bases de datos grandes, en cambio, hay que prever la posibilidad de tener que restringir grandes conjuntos. Habrá que poner énfasis en las técnicas que permiten afinar búsquedas a partir de los resultados previos.

Con el fin de aumentar la recuperación, además de aconsejar que se empleen términos alternativos, puede ser útil ofrecer un diccionario de los que están contenidos en el texto, o bien un tesaurus, ya que aún cuando no se utilicen los descriptores en la búsqueda, las palabras que los integran o aparecen en las relaciones pueden ser de mucha ayuda, sobre todo para usuarios avanzados. La presentación de tesaurus presenta el riesgo de incentivar el uso de expresiones compuestas en sistemas que no lo permiten.

La posibilidad de ajustar la estrategia planteada basándose en los resultados parciales es también de crucial importancia. Como ya se ha expuesto, las búsquedas temáticas avanzan desde una situación de ignorancia hacia una de mayor conocimiento; la mayor parte de los elementos que permiten ese progreso provienen del análisis de los resultados parciales y de las modificaciones que se introducen en consecuencia. Esta actividad puede ser favorecida por la posibilidad de trabajar con conjuntos formados previamente para restringirlos, ampliarlos o modificarlos de cualquier manera. Sostener la continuidad de este proceso exploratorio es una condición para que muchas búsquedas puedan llegar a buen término (Kuhlthau, 1991).

## **Conclusión**

La indización basada en la lengua natural ofrece múltiples posibilidades de recuperación tanto a especialistas como a usuarios finales. Debido a sus características puede ser empleada por estos últimos en múltiples circunstancias, en especial cuando no se pretende efectuar búsquedas temáticas exhaustivas, aún cuando no tengan un entrenamiento específico. Sin embargo, su aprovechamiento presenta algunas dificultades.

Quienes diseñan interfaces deben asumir la tarea de preparar ayudas que aporten de manera ágil soluciones a los principales problemas que encuentran los usuarios, resistiendo la tendencia de estructurarlas con el objetivo de responder únicamente a planteos propios de especialistas en búsquedas. En efecto, son escasas las personas dispuestas a profundizar en los detalles de un lenguaje de recuperación. Para atender a las necesidades de todos conviene poner ejemplos de uso de la interfaz en un lugar de fácil acceso, así como ofrecer una introducción a las principales características del sistema, considerando en especial los

operadores básicos. Para los usuarios que desean perfeccionar sus habilidades o deben solucionar problemas concretos, serán útiles las guías basadas en acciones usuales y la documentación sobre características avanzadas de la interfaz.

## **Bibliografía**

AltaVista. 1999. AltaVista help <<http://www.altavista.com/av/content/help.htm>> [Consulta 29 setiembre 1999].

Bates, Marcia J. 1996. Human, database, and domain factors in content indexing and access to digital libraries and the Internet. <<http://edfu.lis.uiuc.edu/allerton/96/s1/bates.html>> [Consulta: 15 setiembre 1999] En Allerton Institute (38<sup>th</sup>: 1996: Allerton) <<http://edfu.lis.uiuc.edu/allerton/96/>> [Consulta: 15 setiembre 1999].

Brooks, Terrence A. 1998. Orthography as a fundamental impediment to online information retrieval. En *Journal of the American Society for Information Science*. Vol.49, no. 8, 731-741.

Carnegie Mellon University. Lycos. 1999. Lycos help. <<http://www.lycos.com/help/>> [Consulta 29 setiembre 1999].

Collantes, Lourdes Y. 1995. Degree of agreement in naming objects and concepts for information retrieval. En *Journal of the American Society for Information Science*. Vol. 46, no. 2, 116-132.

Connell, Tschera Harkness. 1995. Subject searching in online catalogs: metaknowledge used by experienced searchers. En *Journal of the American Society for Information Science*. Vol. 46, no. 7, 506-518.

Deharbe, Hipólito. 1999. Estrategias de búsqueda por texto secuencial (texto libre). En Comisión Nacional de Energía Atómica. Centro de Información CAC. Manual para instructores de Winisis. Buenos Aires: CNEA. p. 146-155.

Dialog Corporation. 1998. Search basics using DialogWeb. Mountain View: Dialog Corporation. Pag. var.

Iivonen, Mirja y Diane H. Sonnenwald. 1998. From translation to navigation of different discourses: a model of search term selection during the pre-online stage of the search process. En *Journal of the American Society for Information Science*. Vol.49, no. 4, 312-326.

Kuhlthau, Carol C. 1991. Inside the search process: information seeking from the user's perspective. En *Journal of the American Society for Information Science*. Vol.42, no. 5, 361-371.