

# Capítulo1. Introducción

## Evolución y características de las tecnologías de la información y comunicación

Desde los orígenes de la historia de la humanidad, las civilizaciones se transformaban y se estructuraban gracias a los logros conseguidos a partir del aprovechamiento de una tecnología determinada. El impacto del uso del fuego, carbón o petróleo en la producción industrial fue sustituido por el descubrimiento de nuevos y mejores materiales, tales como el plástico o los superconductores. Como es conocido, en la sociedad que surgió tras la revolución industrial el recurso básico era la energía. Para mejorar la efectividad de la producción, el objetivo fundamental fue, extender y ampliar la fuerza del cuerpo humano. De este modo se inventaron máquinas que ahorran trabajo físico y hoy día gran parte de los hombres y mujeres del mundo desarrollado se ha liberado de la inmensa mayoría de difíciles tareas manuales que tenían a finales del siglo XIX.

Actualmente nuestra sociedad, a diferencia de otras anteriores, ya no se articula exclusivamente alrededor de una fuente energética o un conjunto de materias primas, sino que se encuentra protagonizando un proceso vertiginoso en el que todo comienza a estructurarse a partir de un serie de avances técnicos de naturaleza más etérea, por así decirlo, que se ha convenido en denominar "*tecnologías de la información y comunicación*".

La denominación de estas tecnologías indica que éstas están destinadas al manejo y el uso de la información. Desde mediados del siglo XX, se ha reconocido explícitamente que la información es un recurso, y que el desarrollo de herramientas cada vez más eficientes para su manejo está ocasionando un enorme impacto en la productividad. Esto adquiere un sentido especial en esta nueva etapa en que la sociedad emplea cada vez con mayor integración este tipo de tecnologías, dado el hecho de que casi todas las funciones sociales básicas están fuertemente influenciadas por los archivos electrónicos, el teleprocesamiento de la información y las tecnologías de procesamiento del conocimiento. Y por ellos, es que, la información se sitúa en el centro de la vida productiva, convirtiéndose en un recurso estratégico, y, en consecuencia, el objetivo de la sociedad post-industrial se torna ahora en extender y ampliar la capacidad de tomar decisiones acertadas y lo más rápido posible.

En los últimos años se han producido notables adelantos en materia de estas tecnologías, gracias a los cuales cada nueva generación de equipos y programas computacionales ha suscitado rápidamente innovaciones y posibilidades importantes de convergencia con otras tecnologías. Han surgido combinaciones imprevistas de medios tecnológicos de alcance mundial que han configurado la llamada "sociedad de la información" y que representan un desafío para todos los que viven en ella. La evolución hacia esta sociedad y sus efectos económicos,

sociales, culturales y educacionales asociados están provocando cambios considerables en las formas de dirección, creatividad, cooperación, intercambio y compartición de ideas y de conocimiento y, en general, en la vida diaria.

Se pueden destacar **tres características fundamentales** que manifiestan los desarrollos tecnológicos, a saber:

En primer lugar, su gran capacidad para almacenar enormes cantidades de información.

En segundo lugar, sus nuevas fórmulas de comunicación entre los individuos, rompiendo los obstáculos espaciales y temporales, que en otra época eran impensables. Con esto nos referimos a las redes de comunicación y a la telemática, a su importancia para la transmisión de la información y a su capacidad de desarrollar trabajos colaborativos entre diferentes entidades y personas.

En tercer lugar, la capacidad de tratamiento de la información, y no sólo textual, sino de otros sistemas de símbolos, hoy muy arraigados en la vida del ciudadano como son los lenguajes audiovisuales, multimedia, hipertexto; etc.

Desde la invención de la imprenta no se había producido un fenómeno social y tecnológico tan importante en el área de la comunicación como el invento de Internet. El dinamismo de este fenómeno está cambiando los esquemas socioeconómicos preestablecidos y comienza a influir también en todos los ámbitos de la sociedad; es obvio que la revolución tecnológica ha dado lugar a cambios en roles que desempeñan los profesionales relacionados con el manejo y el uso de la información.

Como es conocido, las instituciones de la información han sido creadas para asegurar la accesibilidad a largo plazo de la información registrada. Eso es lo que hacen ahora, y eso es lo que harán en el futuro. Ello significa que adquieren, catalogan o procesan, o sea **organizan y ofrecen** para el uso y **preservan** el material disponible al público, independientemente de la forma en que éste se halle recogido, de tal manera que, cuando sea necesario, pueda ser localizado y utilizado. Esta es la función distintiva de estas instituciones y ninguna otra entidad lleva a cabo este trabajo sistemático a largo plazo. Nada de eso cambia en el mundo digital. La biblioteca como una de las instituciones más antiguas que el hombre creó para preservar sus conocimientos (además del archivo), ha sufrido en sí misma la repercusión de los cambios que a lo largo de la humanidad han experimentado los soportes en los cuales se han plasmado dichos conocimientos. Así, ha tenido que adaptar sus procesos a las condiciones que ha impuesto el surgimiento de nuevos medios creados por el desarrollo tecnológico. Los especialistas en bibliotecología y ciencia de la información deben estar preparados para sacar provecho de estos cambios y proponer los que aún han de mejorar los sistemas actuales, adelantándose a las necesidades de sus usuarios.

## Bibliotecas Digitales. Su presente y futuro

Esta nueva realidad impone la exigencia de desarrollar nuevos métodos de búsqueda, introducir los sistemas que tratan de agregar información semántica a los documentos, nuevas formas de indización y nuevos servicios. Por otra parte, la necesidad facilitar el uso de las herramientas computacionales desencadena una amplia actividad investigadora en torno a interfaces de usuario.

En este contexto, no cabe lugar a dudas que las bibliotecas y sus profesionales deben convertirse de discretos transmisores de información, en activos agentes de cambio. Para lo cual se hace necesario contar con fondos bibliográficos automatizados y también una incorporación dinámica de la nueva tecnología computacional sin la cual no es posible hoy en día hablar de información actualizada, ya que no se pueden ignorar las facilidades que brindan las NTIC para la reproducción y difusión de la información documentaria, lo que permite acelerar la circulación de ésta. La necesidad de encontrar información pertinente, precisa y en el momento oportuno, en tal volumen de documentos, que además son heterogéneos, provoca que los sistemas de recuperación de información estén en un momento crucial de cambio.

Por ello, que en estos momentos podemos hablar de una biblioteca en proceso de transformación tecnológica que, partiendo de la **información** y los servicios ya existentes, busca la integración entre la biblioteca tradicional y la digital. Existen numerosas reflexiones sobre el futuro de las **bibliotecas**, y en casi todas ellas podemos encontrar recogido implícitamente los conceptos de **integración, sistemas híbridos, evolución y convivencia**.

Teniendo en cuenta lo expresado hasta aquí, se pueden mencionar las siguientes características atribuidas a las **Bibliotecas Digitales**.

- Reproducen su *información primaria* en formato digital
- Prestan servicios simultáneamente
- Generan servicios distribuidos
- No tienen límite de horario
- Cuentan con fuentes de información muy actualizadas, completas y variadas
- Permiten compartir recursos
- Contienen diferentes formatos para el manejo de la información
- Brindan a los usuarios la oportunidad de publicar su información y compartirla con otros usuarios
- Sus servicios se complementan con el **uso de multimedia** y con otras herramientas como la **videoconferencia**.

La biblioteca digital se concibe a partir del uso intensivo de las tecnologías de la información y de la comunicación (TIC) en su diseño y formas de servicios. No se trata únicamente de cambiar la forma y maneras de difundir la información, sino de

ensayar nuevas formas de conservar, organizar y difundir el conocimiento por medio de colecciones de documentos digitales.

Este tipo de bibliotecas se pueden considerar como uno de los sistemas de información más novedosos y complejos que brinda a los usuarios un espacio muy amplio para la colaboración; sus servicios se complementan con el **uso de multimedia** y con otras herramientas como la **videoconferencia**.

El acceso al conocimiento y a la cultura desde una variedad de fuentes (tanto tradicionales como de las más recientes, en términos de formato y uso), en una variedad de lenguas, visiones y tradiciones incluye el proceso de transformar ese conocimiento a un formato accesible. Es evidente que estos problemas deberían tratarse dentro del contexto de la biblioteca digital, que debe afianzar cada día más su papel clave en la emergente sociedad de la información y el conocimiento como un lugar que asegure la preservación y la migración de información de tal manera que pueda ser utilizada por todos los que la necesiten.

A partir de lo expuesto anteriormente, se puede a concluir que el nuevo profesional de la información debe ser formado en cuestiones de diversa índole: empresariales (conocimiento en el medio que se mueve, bien en un organización concreta, bien en función del sector en el que se sitúa), relativas a la información (conocimientos de los diferentes discursos, análisis semánticos, lógica, sistemas de recuperación) y tecnológicas (inteligencia artificial, software, redes). De ahí también es la importancia del estudio y el conocimiento de los procesos de **Búsqueda y Recuperación de la Información**.

## **Capítulo 2. Proceso de búsqueda y Recuperación de la Información**

### **El proceso genérico de búsqueda de información. Tipos de búsqueda**

La búsqueda y recuperación de la información entre grandes volúmenes de datos es un tema que se hace cada día más relevante, en un mundo que vive en la super-abundancia de la producción de información. Un ejemplo muy cercano es Internet, hipermedio en el que diariamente se crean miles de páginas de los temas más diversos, y que se suman a la producción de más de una década de documentos hipertextuales.

La búsqueda de información en entornos informáticos es una tarea compleja, por cuanto no se desarrolla de una manera uniforme, y está sujeta a variables como los conocimientos previos, así como las expectativas y las estrategias de quienes buscan la información. Además, en este proceso de búsqueda, las herramientas

con que se realice la tarea y las capacidades de dichas herramientas para proveer de resultados, son también fundamentales.

Pocas veces la búsqueda de información es desarrollada por se, probablemente porque la información es un hito en el proceso de producción y elaboración de datos, tendientes al conocimiento y la acción. En los estudios acerca de las tareas de búsqueda de información, se evidencia que la búsqueda de información es sólo una pequeña porción del proceso de trabajo informativo, cuyo mayor tiempo es utilizado en la lectura, análisis y resumen de la información recuperada. En la misma dirección, se plantea también que el acceso a la información es un proceso en el cual la búsqueda y recuperación de datos es solamente una parte menor, observando que la mayor parte del trabajo consiste en dotar de sentido a la información, con miras a resolver un problema.

No obstante, el proceso de búsqueda y recuperación de información, es de vital importancia en el desarrollo de conocimiento, pues en esta etapa se obtiene el material que permite reducir la incertidumbre y ampliar los elementos de juicio para la toma de decisiones y la resolución de problemas.

Al iniciar un proceso de búsqueda de información, las personas tienen distintas expectativas y metas, y utilizan los sistemas de búsqueda como una herramienta para ayudar a alcanzar esos objetivos. Sus necesidades pueden variar desde información básica, hasta detallados y profundos informes sobre uno o varios temas; por ejemplo, desde encontrar una dirección a mantenerse informado sobre los avances en el tratamiento de una enfermedad. Considerando las diversas expectativas de información, Rosenfeld identifica cuatro tipos de búsqueda:

*“Búsqueda de elemento conocido: Consiste en una búsqueda acotada. Cuenta con preguntas (queries) bien definidas y con pocas alternativas de (o sólo una) respuesta posible. Habitualmente se conoce dónde buscar y es el tipo de búsqueda más sencillo.*

*“Búsqueda de existencia: En este tipo de búsqueda existe una respuesta clara, pero se desconoce cómo realizar adecuadamente la pregunta. Se trata habitualmente de una búsqueda de información en que el valor del contexto es muy importante y suele tratarse de temas complejos que articulan distintas variables o de temas poco conocidos.*

*“Búsqueda exploratoria: Se trata de una búsqueda en la que se desconoce exactamente lo que se quiere encontrar. En este caso el tema de búsqueda es amplio y por ende no existe una pregunta definida. Habitualmente, la expectativa es conocer algo en términos generales o introducirse en un tema.*

*“Búsqueda Global: El objetivo de esta búsqueda es conocer con profundidad un tema. Se trata de inspeccionar todos los aspectos y detalles de una materia, para lo cual habrá que formular la pregunta combinando variables en las que la query incluya términos asociados”. (Rosenfeld, L., 1998)*

Estos tipos de búsqueda no son estáticos y, como señala el propio Rosenfeld, durante el proceso de búsqueda podrán variar las necesidades de información requeridas por el usuario, con lo cual el tipo de búsqueda realizada cambiará de manera correspondiente. Además, la búsqueda de información es iterativa, es decir, los usuarios aprenden algo con cada respuesta del sistema y con cada refinamiento de la pregunta.

Por otra parte, resulta fundamental considerar que las habilidades de quienes realizan la búsqueda de información, son determinantes en la consecución de resultados. Estas habilidades deben ser ponderadas en dos ámbitos: el primero, ya esbozado en la caracterización de los tipos de búsqueda, se refiere al conocimiento que pueda tener el usuario respecto de un tema; el segundo se refiere a las habilidades de los sujetos para usar determinados sistemas de búsqueda.

El conocimiento de un tema, en tanto más acabado, posibilita la identificación de la información precisa requerida para la solución de un problema, así como la adecuada identificación de las fuentes para la selección de información relevante. Este no es problema menor, *“por cuanto nos sitúa además, en el eterno dilema semántico: la nomenclatura, la categorización y de la indización de contenidos responden más a una cuestión de prácticas lingüísticas que a una lógica universal”*. (Quillian, R, 1998)

En cuanto al manejo de sistemas de búsqueda de información, deben considerarse al menos dos situaciones extremas, la de aquellas personas que cuentan con un acabado conocimiento del sistema y la de aquellos que son usuarios novatos y carecen del aprendizaje que les permita utilizar de manera automática los recursos que la herramienta de búsqueda ofrece. *“En ambos casos se trata de que la herramienta resulte invisible, es decir, que los usuarios no requieran amplios recursos cognitivos para el manejo del sistema, utilizando un tipo de memoria llamada procedural y reservando recursos atencionales y de razonamiento para el contenido en sí”*. (Norman, 1990).

Con independencia del grado de conocimiento sobre un tema o habilidad para el manejo de un sistema de búsqueda, y más allá del tipo de búsqueda realizada, - desde revisar un tema conocido hasta explorar un tema sin referencias - se ha definido un modelo para el proceso de acceso a la información. En este modelo se plantea que existe una serie de interacciones entre el usuario y el sistema de búsqueda, las que han sido caracterizadas por Schneiderman en un proceso estándar que se puede describir en la siguiente secuencia de pasos:

1. *Comienza con una necesidad de información.*
2. *Selección de un sistema y colecciones de datos para buscar información.*
3. *Formulación de una consulta.*
4. *Envío de la consulta (query) al sistema.*
5. *Recepción de los resultados en la forma de documentos.*
6. *Exploración, evaluación, e interpretación de los resultados.*

7. *Detención, o,*
8. *Reformulación de la consulta y vuelta al paso 4 (Schneiderman, 1997)*

Otro autor, estudiioso en este campo, Herst ha realizado un análisis crítico del modelo y plantea una serie de dificultades que deben ser consideradas y resueltas al diseñar sistemas de búsqueda. Estas dificultades del modelo pueden ser resumidas desde su punto de vista de la manera siguiente:

- *Demando del usuario un acabado conocimiento en la formulación de la consulta, de lo contrario puede obtener una lista larga (y desorganizada) de resultados, o resultados vacíos.*
- *Asume que la necesidad de información del usuario es estática y desconoce que el proceso de búsqueda de información es iterativo.*
- *Desconoce que los usuarios aprenden durante el proceso de la búsqueda. Al explorar los resultados (leer títulos de resultados, leer documentos recuperados, ver listas de temas relacionados y navegar dentro de los sitios vinculados) se produce aprendizaje.*
- *La exploración y la navegación dentro de una búsqueda inexacta también puede generar un conocimiento por descubrimiento.*
- *Minimiza la posibilidad de retroalimentación de relevancia, que conduciría a explorar términos asociados, o las descripciones temáticas de las colecciones de documentos.*
- *Minimiza el papel de la selección del recurso, que es cada vez más importante ahora que, miles de colecciones de datos son inmediatamente accesibles.*

(Herst, 1999)

Este autor señala que el Modelo Cosecha de Bayas o Recolección de frutos, sería mejor ajustado a las actuales posibilidades de búsqueda de información. El modelo mencionado plantea que las necesidades de información de los usuarios, y por lo tanto sus consultas, cambian continuamente, esto sucede debido a que la información encontrada puede conducir a una nueva e inesperada dirección, disminuyendo la prioridad de una meta en favor de otra. Además, las necesidades de información de los usuarios no son satisfechas por una recuperación final de documentos, ni por el refinamiento dentro de un conjunto de documentos, sino por una serie de selecciones parciales dentro de documentos de distinto origen y por información fragmentada encontrada a lo largo de búsqueda.

Por último, Herst subraya, que el principal valor de la búsqueda reside en el aprendizaje y la adquisición de la información acumulados durante la búsqueda, más que en los resultados finales y, finalmente, apunta que el contexto del problema se mantiene y es llevado de una etapa de búsqueda a la siguiente.

## **Bibliografía Citada.**

HERST, M. " User Interfaces and Visualization". En Modern Information Retrieval. Baeza-Yates, Ribeiro-Neto. Addison-Wesley-Longman Publishing. EE.UU., 1999.

NORMAN, D. " Psicología de los Objetos Cotidianos". Editorial Narea. España, 1990.

QUILLIAN, R "Semantic Memory". En Readings in Cognitive Science: A Perspective from Psychology & Artificial Intelligence. Editado por Collins, A y Smith, E. Ed. Morgan Kaufmann. EE.UU, 1988.

ROSENFELD, L. "Arquitectura de la información ". McGraw Hill-O'Reilly. España, 1998.

SCHNEIDERMAN, B. "Clarifying Search: A User-Interface Framework for Text Searches". Department of Computer Science and Human-Computer Interaction Laboratory University of Maryland. 1997.

En <http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html>

## On conceptual models for information seeking and retrieval research

**Kalervo Järvelin**

**Centre for Advanced Study**

**University of Tampere, Tampere, Finland**

**and**

**T.D. Wilson**

**Visiting Professor**

**Högskolan i Borås, Borås, Sweden**

### Abstract

*There are several kinds of conceptual models for information seeking and retrieval (IS&R). The paper suggests that some models are of a summary type and others more analytic. Such models serve different research purposes. The purpose of this paper is to discuss the functions of conceptual models in scientific research, in IS&R research in particular. What kind of models are there and in what ways may they help the investigators? What kinds of models are needed for various purposes? In particular, we are looking for models that provide guidance in setting research questions, and formulation of hypotheses. As a example, the paper discusses [at length] one analytical model of task-based information seeking and its contribution to the development of the research area.*

### Introduction

There has been considerable interest in recent years in producing conceptual models for information seeking and retrieval (IS&R) research. The recent paper by Wilson (1999) reviews models for information behaviour (Wilson 1981), information seeking behaviour (Wilson 1981; 1996; Dervin, 1986; Ellis et al. 1993, Kuhlthau, 1991), and information searching or retrieval (Ingwersen, 1996; Belkin, et al. 1995; Spink, 1997).

Wilson (1999: 250) notes concerning the models of information behaviour, among others, that "rarely do such models advance to the stage of specifying relationships among theoretical propositions: rather they are at a pre-theoretical stage, but may suggest relationships that might be fruitful to explore or test." Later he notes that,

*"[t]he limitation of this kind of model, however, is that it does little more than provide a map of the area and draw attention to gaps in research: it provides no suggestion of causative factors in information behaviour and, consequently, it does not directly suggest hypotheses to be tested." (1999: 251)*

It seems, therefore, that there may be several kinds of conceptual models for IS&R and that, at least for some research purposes, we would need models that may suggest relationships that might be fruitful to explore and provide hypotheses to test. The purpose of this paper is to discuss the functions of conceptual models in scientific research, in IS&R research in particular. What kind of models are there and in what ways may they help the investigators? What kinds of models are needed for various purposes? In particular, we are looking for models that provide guidance in setting research questions, and formulating hypotheses.

In the following section we shall discuss the meaning and function of conceptual frameworks and principles for judging their merits in research. This extends Järvelin's (1987) discussion on criteria for assessing conceptual models for IS&R research. Section 3 analyses briefly some summary frameworks for the IS&R domain. This is followed by a discussion of analytic frameworks. In particular, the classifications suggested by Järvelin are presented and their use in generating fruitful research hypotheses is discussed. Jarvelin's suggestions led to empirical study (Byström & Järvelin, 1995; Byström, 1999) and theoretical development (Byström, 1999; Vakkari & Kuokkanen, 1987; Vakkari, 1999), which analysed the relationships of task complexity and information seeking. The uses for the classifications in later research are briefly summarised. The paper ends with discussion and conclusions.

### ***Conceptual models and their uses***

All research has an underlying model of the phenomena it investigates, be it tacitly assumed or explicit. Such models, called conceptual frameworks (Engelbart, 1962) or conceptual models, easily become topics of discussion and debate when a research area is in transition. Often two or more models are compared and debated. With an eye on advancing the research area, how should the models be assessed for their possible uses? In this section we discuss the function of conceptual frameworks and principles for judging their merits.

According to Engelbart, developing conceptual models means specifying the following:

- Essential objects or components of the system to be studied.
- The relationships of the objects that are recognised.
- What kinds of changes in the objects or their relationships affect the functioning of the system - and in what ways.
- Promising or fruitful goals and methods of research.

Conceptual models are broader and more fundamental than scientific theories in that they set the preconditions of theory formulation. In fact, they provide the conceptual and methodological tools for formulating hypotheses and theories. If they are also seen to represent schools of thought, chronological continuity, or principles, beliefs and values of the research community, they become paradigms.

The conceptual model of a research area is always constructed - it does not simply lie somewhere waiting for someone to pick it up.

The literature of the Philosophy of Science provides discussions on the functions of scientific theories. According to Bunge (1967), scientific theories are needed (or used) for the following functions:

- Systematisation of knowledge by:
  - Integrating formerly separate parts of knowledge.
  - Generalising and explaining lower abstraction level knowledge (or observations, data) through higher level constructs.
  - Explanation of facts through systems of hypotheses which entail the facts.
  - Expanding knowledge by deducing new propositions based on selected starting points and collected information.
  - Improving the testability of hypotheses through the control context provided by systems of hypotheses.
- Guiding research by:
  - Pointing to fruitful problems.
  - Proposing the collection of data, which nobody would understand to collect without the theory.
  - Proposing totally new lines of research.
- Mapping a portion of reality by:
  - Representing or modelling the objects (and relationships) of that chunk instead of just summarising the data.
  - Providing a tool for producing new data.

We believe that these functions are also suitable functions of conceptual models, which are more general in nature than theories. Clearly, conceptual models may and should map reality, guide research and systematise knowledge, for example, by integration and by proposing systems of hypotheses.

A conceptual model provides a working strategy, a scheme containing general, major concepts and their interrelations. It orients research towards specific sets of research questions. A conceptual model cannot be assessed directly empirically, because it forms the basis of formulating empirically testable research questions and hypotheses. It can only be assessed in terms of its instrumental and heuristic value. Typically, this happens by assessing the research strategies and programmes (and results) it creates. The latter programmes consist of interrelated substantial theories and research relevant for evaluating them (Wagner, et al., 1992; Vakkari 1998). If the substantial theories prove to be fertile, the model is so too.

However, waiting for the substantial theories to prove to their fertility may take some time. In the meantime, or even before embarking on some line of research, it may be important to argue about the merits of various conceptual models. The

following are the types of arguments that can be used to judge the merits of a conceptual model:

- General scientific principles:
  - When studying some phenomena, they should be studied in all situations, and also under extreme conditions (cf. thermophysics). Thus, you do not just consider information seeking by academics but also by other professions or by laymen.
  - The framework should be limited in a meaningful way as a system. For understanding information seeking by human actors, the proper system is not some service (for example, a library) and its clients but rather an information actor immersed in his or her situation and information environment (for example, all information access systems).

When two competing conceptual models are compared the following criteria may be applied to judge their merits:

- Simplicity: simpler is better other things being equal.
- Accuracy: accuracy and explicitness in concepts is desirable.
- Scope: a broader scope is better because it subsumes narrower ones, other things being equal.
- Systematic power: the ability to organise concepts, relationships and data in meaningful systematic ways is desirable.
- Explanatory power: the ability to explain and predict phenomena is desirable.
- Reliability: the ability, within the range of the model, to provide valid representations across the full range of possible situations.
- Validity: the ability to provide valid representations and findings is desirable.
- Fruitfulness: the ability to suggest problems for solving and hypotheses for testing is desirable.

Theoretical development or the construction of new conceptual models in any research area often requires conceptual and terminological development. Conceptual development may mean fulfilling, perhaps in a better way than before, the basic requirements for scientific concepts - precision, accuracy, simplicity, generality, and suitability for expressing propositions, which may be shown true or false. Moreover, good concepts represent essential features (objects, relationships, events) of the research area. More importantly, the concepts should differentiate and classify the phenomena in ways that lead to interesting hypotheses (or research problems). This means that the concepts must relate to each other in systematic and fruitful ways. Concepts also need to support research into the phenomena by known research methods (or, somewhat relaxed, by methods that can be developed). They need to be compatible with each other and with research methods (that is, be congruent).

## **Summary frameworks**

### **Two sample frameworks**

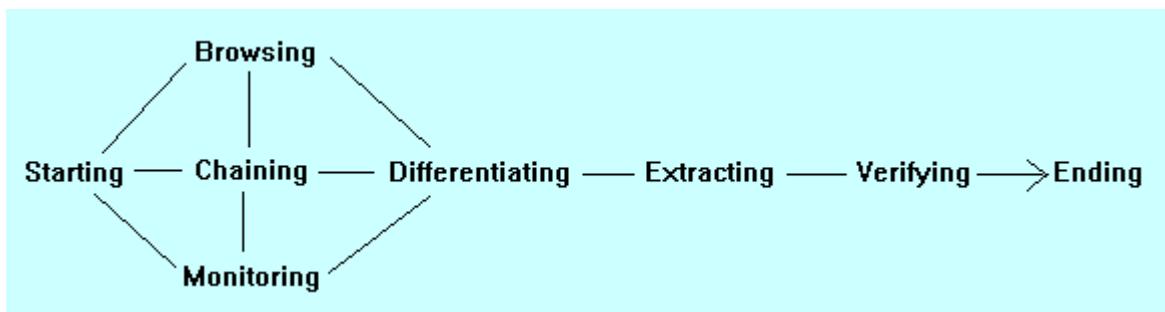
We will discuss Ellis's (1989; Ellis, et al., 1993) and Ingwersen's (1996) frameworks. These are used and discussed here as examples only and we make no claims about their merits with respect to the research tasks for which they were originally intended.

Ellis's elaboration of the different behaviours involved in information seeking consists of six features. Ellis makes no claims to the effect that the different behaviours constitute a single set of stages; indeed, he uses the term 'features' rather than 'stages'. These features are named and defined below:

- Starting: the means employed by the user to begin seeking information, for example, asking some knowledgeable colleague;
- Chaining: following footnotes and citations in known material or 'forward' chaining from known items through citation indexes;
- Browsing: 'semi-directed or semi-structured searching' (Ellis, 1989: 187);
- Differentiating: using known differences in information sources as a way of filtering the amount of information obtained;
- Monitoring: keeping up-to-date or current awareness searching;
- Extracting: selectively identifying relevant material in an information source;
- Verifying: checking the accuracy of information;
- Ending: which may be defined as 'tying up loose ends' through a final search.

The strength of Ellis's model is that it is based on empirical research and has been tested in subsequent studies, most recently in the context of an engineering company (Ellis & Haugan, 1997).

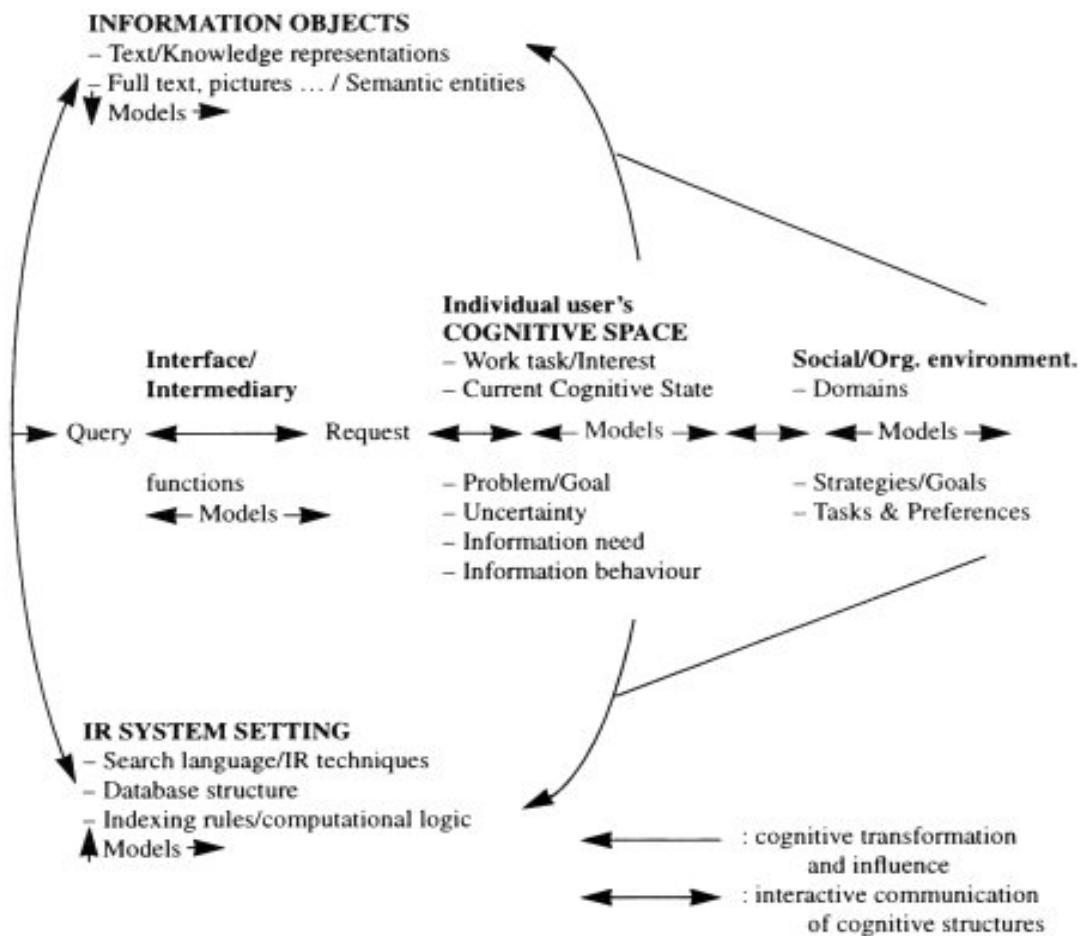
Of the features, Ellis (1989: 178) notes that, '*...the detailed interrelation or interaction of the features in any individual information seeking pattern will depend on the unique circumstances of the information seeking activities of the person concerned at that particular point in time*'. Wilson (1999) proposes how these features may relate to each other temporally, providing a partial order; see Figure 1.



**Figure 1. A process version of Ellis's behavioural framework (Wilson 1999)**

One may *describe* any information seeking activities through Ellis's features. Indeed, they are general enough to fit a large number of empirical situations. However, if one is to *explain* information seeking behaviour, say, in terms of the work tasks the subjects are engaged with, or their knowledge on the task, the features fall short because they are not explicitly related to such external possible causative factors.

Of course, Ellis's model may still be of indirect help in finding explanations for information seeking behaviour. It is possible to discern differences in any of the 'features' in different situations, involving different kinds of persons through successive research projects. For example, some persons in some roles may be shown to engage more or less in monitoring than other persons. This may then lead to an examination of the factors that 'cause' these differences.



**Figure 2. Ingwersen's model of the IR process (Wilson, 1999; based on Ingwersen, 1996)**

Ingwersen's (1996) model is slightly simplified in Figure 2. Wilson points out its relationships to other models of information seeking behaviour. In particular, the elements *user's cognitive space* and *social/organisational environment*, resemble the *person in context* and *environmental factors* specified in Wilson's models (1981, 1996; 1999). The general orientation towards queries posed to an IR system point to a concern with the *active search*, which is the concern of most information-seeking models. Ingwersen, however, makes explicit a number of other elements: first, he demonstrates that within each area of his model, the functions of the information user, the document author, the intermediary, the interface and the IR system are the result of explicit or implicit cognitive models of the domain of interest at that particular point. Thus, users have models of their work-task or their information need, or their problem or goal, which are usually implicit, but often capable of explication. Again, the IR system is an explication of the system designer's cognitive model of what the system should do and how it should function. Secondly, Ingwersen brings the IR system into the picture, suggesting that a comprehensive model of information-seeking behaviour must include the system that points to the information objects that may be of interest to the enquirer. Thirdly, he shows that various cognitive transformations take place in moving from the life-world in which the user experiences a problem or identifies a goal to a situation in which a store of pointers to information objects can be satisfactorily searched and useful objects identified. Finally he points to the need for these cognitive structures and their transformations to be effectively communicated throughout the 'system', which will include the user, the author and the IR system designer. All this true—it is easy to agree.

Thus, Ingwersen's model, to a degree, integrates ideas relating to information behaviour and information needs with issues of IR system design, and this is an important strength of the model. Saracevic suggests that (1996): '*The weakness is in that it does not provide for testability... and even less for application to evaluation of IR systems.*' However, recently, Borlund and Ingwersen (1997; 1998; Borlund, 2000) have developed and tested an evaluative strategy on the basis of this model and have demonstrated its value in testing interactive IR systems. A remaining potential weakness is that information behaviour other than information retrieval is not explicitly analysed. Issues of how users arrive at the point of making a search, and how their cognitive structures are affected by the processes of deciding how and when to move towards information searching, may be lost. These issues may be discussed in terms of the social or organisational environment but, to say the least, this is not explicit.

In Ingwersen's model, there are several entities of the IS&R interplay present, and some of their relevant features are explicated. Therefore, there are better possibilities for formulating research questions for empirical study; for example, how is an individual user's uncertainty related to the intermediary functions, and how does this affect the retrieval process? However, there is still some way to go before one may say that an empirical research problem has been specified. This could be done by classifying, for example, uncertainty and intermediary functions in ways that suggest empirical relationships.



## Uses of summary frameworks

Summary models provide overviews of research domains, and list factors affecting the phenomena. It is often easy to agree that, what the models propose, are factors affecting the processes of interest. However, without detailed analysis of the components, such models provide little or no suggestion of causative factors in IS&R phenomena and, consequently, they do not directly suggest hypotheses to be tested. Indirectly, however, a comparison of findings across several studies may suggest causative factors to be explored.

### ***An analytic framework***

Järvelin (1987) suggested three classifications and discussed their use in generating fruitful research hypotheses for the analysis of the relationships of task complexity and information seeking. Byström and Järvelin (1995; Byström 1999) revised the classification and carried out an empirical study, and Byström (1999) and Vakkari (1998; 1999) suggested theoretical developments. We first present the classifications and then discuss their theoretical and methodological consequences.

### **Task complexity**

A worker's job consists of tasks, which consist of levels of progressively smaller subtasks. Tasks are either given to, or identified by, the worker. Each task has a recognisable beginning and end, the former containing recognisable stimuli and guidelines concerning goals and/or measures to be taken (Hackman, 1969). Seen in this way, both a large task or any of its (obviously simpler) sub-tasks may be considered as a task. This relativity in definition is necessary in order to analyse tasks of different levels of complexity.

In information seeking we are interested in information-related tasks. These can be seen as perceived (or subjective) tasks or objective tasks. The relationships of objective and perceived tasks have been considered in organisational psychology (Campbell, 1988; Hackman, 1969; Wood, 1986) where task descriptions based on perceived tasks are generally held invalid for many purposes (for example, Roberts & Glick, 1981). However, in information seeking, perceived tasks must be considered because each worker may interpret the same objective task differently (for example, as regards its complexity) and the perceived task always forms the basis for the actual performance of the task and for interpreting information needs and the choice of promising actions for satisfying them.

The literature suggests many task characteristics related to complexity: repetition, analysability, *a priori* determinability, the number of alternative paths of task performance, outcome novelty, number of goals and conflicting dependencies among them, uncertainties between performance and goals, number of inputs, cognitive and skill requirements, as well as the time-varying conditions of task performance (Campbell, 1988; Daft *et al.*, 1988; Fischer, 1979; Fiske & Maddi,

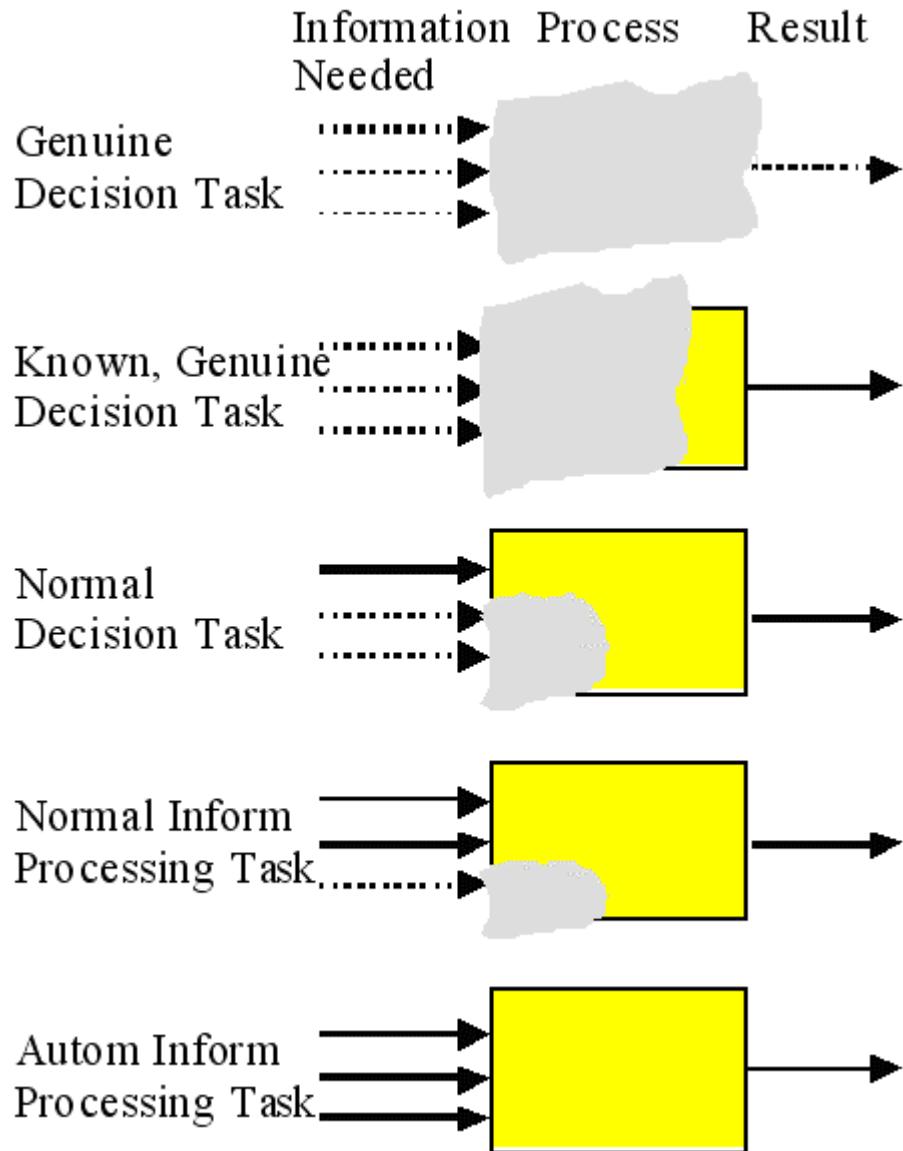
1961; Hart & Rice, 1991; Järvelin, 1986; March & Simon, 1967; MacMullin & Taylor, 1984; Tiamiyu, 1992; Tushman, 1978; Van de Ven & Ferry, 1980; Wood, 1986; Zeffane & Gul, 1993). Also, these characteristics have been understood in many different ways in the literature. They belong in two main groups: characteristics related to the *a priori* determinability of tasks, and characteristics related to the extent of tasks.

Järvelin (1987; Byström and Järvelin, 1995) suggest a simple, one-dimensional categorisation of the complexity of tasks based on, from the worker's point of view, *a priori* determinability of, or uncertainty about, task outcomes, process and information requirements. This dimension is related to the above task characteristics: repetition, analysability, *a priori* determinability, the number of alternative paths of task performance and outcome novelty. Similar one-dimensional categorizations of complexity are used by Tiamiyu (1992) and Van de Ven and Ferry (1980). Simple tasks are routine information processing tasks, where the inputs, process and outcomes can be determined *a priori*, while difficult or complex tasks are new and genuine decision tasks, which cannot be so determined. Such a categorization is generic and, thus, widely applicable to many types of tasks and domains.

### Task categorisation

In this paper, tasks are classified into five categories ranging from an automatic information-processing task to a genuine decision task. This categorization is based on the *a priori* determinability (or structuredness) of tasks and is closely related to task difficulty or complexity.

Task complexity is often seen to depend on the degree of *a priori* uncertainty about the task inputs, process and outcome (for example, Van de Ven & Ferry, 1980). In automatic information processing tasks, the type of the task result, the work process through the task, and the types of information used can all be described in detail in advance. In genuine decision tasks, on the contrary, none can be determined *a priori*.<sup>1</sup> Our task categorization is presented in Fig. 3 where information (both input and result) is represented by arrows and the task process by boxes. The *a priori* determinable parts of tasks are represented by solid arrows and solid boxes, and the *a priori* indeterminable parts of tasks are represented by dashed arrows and shaded boxes. Dashed arrows and shaded boxes thus represent cased-based arbitration. Three arrows are used in the input side to visualise that many inputs often are needed and that there are degrees of *a priori* determinability among them. Also the types of input differ by task category as discussed in the next subsection.



**Figure 3: Task categories (Anon. 1974)**

Tasks in different categories can be characterized briefly as follows:

- *Automatic information processing tasks* are *a priori* completely determinable so that, in principle, they could be automated - whether actually automated or not. Example: computation of a person's net salary yields a real number in some known range and requires this person's gross salary and tax code, and the taxation table.
- *Normal information processing tasks* are almost completely *a priori* determinable but require some case-based arbitration concerning for example, the sufficiency of the information normally collected. Thus part of

the process and information needed is *a priori* indeterminable. Example: tax coding is mostly rule-based but some cases require additional clarification, that is, case-dependent information collection.

- *Normal decision tasks* are still quite structured but in them cased-based arbitration has a major role. Example: hiring an employee or evaluating a student's term paper.
- In *known, genuine decision tasks* the type and structure of the result is *a priori* known but permanent procedures for performing the tasks have not yet emerged. Thus, the process is largely indeterminable and so are its information requirements. Example: deciding about the location for a new factory or medium-range planning in organizations.
- *Genuine decision tasks* are unexpected, new and unstructured. Thus, neither the result, the process nor the information requirements can be characterized in advance. The first concern is task structuring. Example: the collapse of the Soviet Union from the viewpoint of other governments.

Information seeking research has focused mostly on tasks in the middle and upper parts of the categories (normal decision task to genuine decision task) although this dimension has only rarely been recognized. Belkin (1980) describes a similar scale of problem situation levels). The categories above are relative to the worker: what is a genuine decision task to a novice may be a normal decision to an expert.

### **Types of information needed in tasks**

In expert systems design, the types of information are classified as problem information (PI), domain information (DI), and problem solving information (PSI) (for example, Barr & Feigenbaum, 1981). Järvelin and Repo (1983; 1984) proposed these concepts for information seeking research. These information categories can be characterized as follows:

- *Problem information* describes the structure, properties and requirements of the problem at hand. For example, in bridge construction, information on the type and purpose of the bridge and on the building site constitutes problem information. It is typically available in the problem environment, but, in the case of previous problems of the same type, it may also be available in documents.
- *Domain information* consists of known facts, concepts, laws and theories in the domain of the problem. For example, in bridge construction, information on the strength and thermal expansion of steel belongs to domain information. This is, typically, tested scientific and technological information published in journals and textbooks.
- *Problem-solving information* covers the methods of problem treatment. It describes how problems should be seen and formulated, what problem and domain information should be used (and how) in order to solve the problems. For example, in bridge construction, the design engineer's heuristics concerning the pros and cons of various bridge design types

constitute problem-solving information. It is instrumental information and typically available only from knowledgeable persons (or experts).

These three information categories are orthogonal, that is, represent three different dimensions and have different roles in problem treatment. All are necessary in problem treatment but, depending on the task, and to different degrees, may be available to a worker performing the task. Because their typical sources are different, typical channels for acquiring them may also be different.

Regarding Figure 3, the solid arrows representing input information may be seen as *a priori* determinable problem information whereas the dashed arrows would represent all *a priori* indeterminable information, often problem-solving information.

## Types of Information Sources

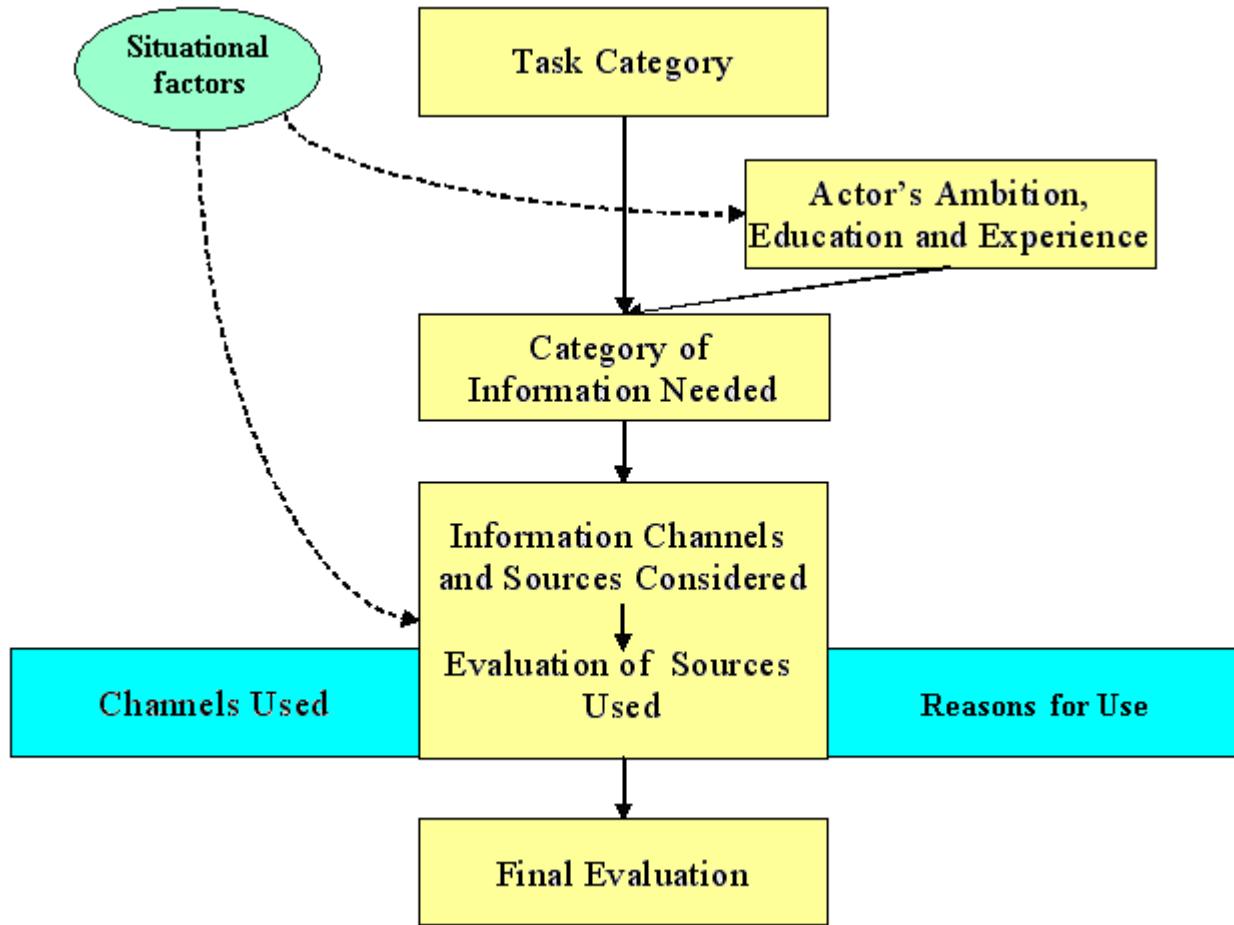
Byström and Järvelin (1995) classified the types of information sources as:

- *fact-oriented*:
  - registers (manual and computerised catalogues and files)
  - commercial databases
- *problem-oriented*:
  - the people concerned (for example, people proposing, or affected by, administrative actions)
  - official documents (for example, agendas, meeting minutes, letters, applications, memoranda, maps, unpublished planning documents)
- *general-purpose*:
  - experts (including knowledgeable colleagues)
  - literature (for example, books, reports, journals, newspapers)
  - personal collections (personal notes, calculations, etc.)

They also classified the sources as being either *internal* or *external* to the organisation in which the user works.

## Theoretical and methodological consequences

Byström and Järvelin used their framework, the three classifications of tasks, information and information sources, for the analysis of their data structured in work charts (Figure 4). In combination, the three classifications suggest a set of hypotheses of the type: "*Tasks of complexity type X require information of type Y that is available from sources of type Z*". Thus the classifications suggest analytical relationships between the variables.



**Figure 4: The work chart structure (Byström and Järvelin, 1995 )**

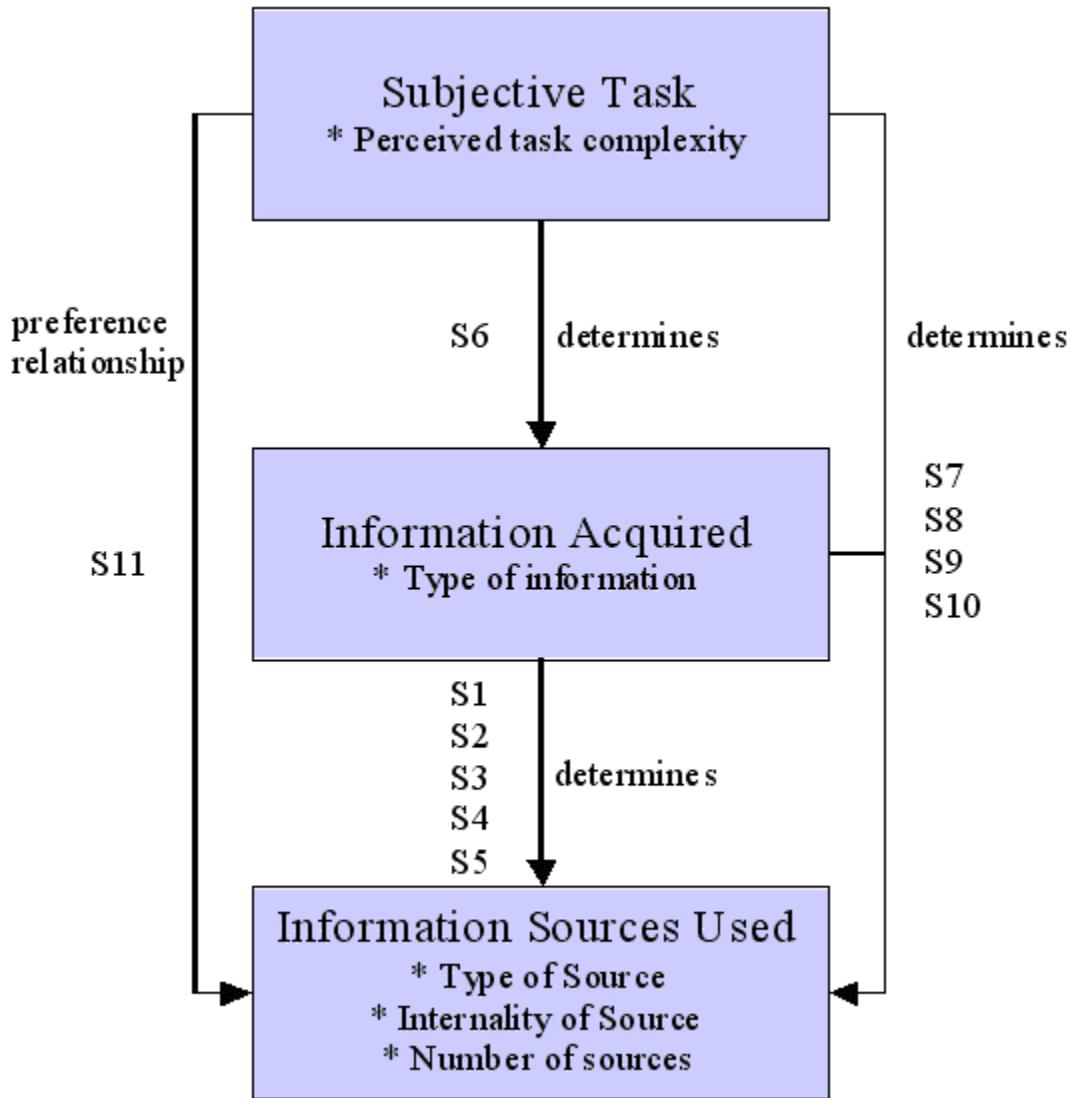
Byström and Järvelin (1995; Murtonen, 1992<sup>2</sup>) developed a qualitative method for task-level analysis of the effects of task complexity on information seeking and found, in a public administration context, that these effects are systematic and logical. The specific research problem studied was: what types of information are sought through which types of channels from what kinds of sources in which kinds of tasks? They found that, as task complexity increased, so:

- the complexity of information needed increased,
- the needs for domain information and problem solving information increased,
- the share of general-purpose sources (experts, literature, personal collections) increased and that of problem and fact-oriented sources decreased,
- the success of information seeking decreased,
- the internality of channels decreased, and
- the number of sources increased.

The contrast between simple and complex tasks underlines the importance and consequences of task complexity: in the latter understanding, sense-making and problem formulation are essential and require different types and more complex information through somewhat different types of channels from different types of sources.

Byström followed on with further empirical studies (1999; Murtonen, 1994). Based on her empirical findings, Byström presented a revised model of task-based information seeking (Figure 5). The model contains eleven statements (S1 - S11 in Figure 5). Some of the statements are given below (all are given in the Appendix):

- S2: the more information types are needed, the greater the share of people as sources.
- S6: the higher the degree of task complexity, the more probable is the need for multiple information types: first task information, then task and domain information, and finally task, domain and [problem] solving information.
- S8: the higher the degree of task complexity, the more information types are needed and the greater the share of general-purpose sources and the smaller the share of task-oriented sources.
- S10: task complexity is distinctly related to increasing internality of people as sources and decreasing internality of documentary sources.
- S11: Increasing task complexity fosters the use of people as sources.



**Figure 5: A model of task-based information seeking (Byström, 1999)**

Vakkari (1998; 1999; Vakkari & Kuokkanen, 1997) analysed, and contributed to, theory growth in task-based information seeking. Vakkari and Kuokkanen apply Wagner & Berger's (1985) analysis of theory growth to reconstructing a theory based on the framework by Byström and Järvelin (1995). Vakkari and Kuokkanen note that the latter did not fully utilise the whole potential of the framework, for example, the relationships of information types and source use was not fully developed. They derive new hypotheses for further empirical work from the reconstructed theory. The resulting theory is thus broader in scope and has more empirical consequences than the original. Vakkari and Kuokkanen state that their reconstruction creates potential growth of knowledge within the theory of information seeking. This is easy to agree.

Vakkari (1998) further uses Wagner & Berger (1985) and focuses on the theoretical research programme starting from Tushman's (1978) study on task complexity and information. He finds that Byström and Järvelin's (1995) work created progress in all dimensions of theory growth, especially in terms of precision and scope. The framework (research programme), by adding the classification of information types, explicated several new factual relations among information seeking phenomena.

The empirical findings and theoretical developments by Byström, Järvelin and Vakkari classify tasks, information and information sources in a systematic way. The latter are also systematically related to other central concepts of information seeking in a systematic way. The original papers suggested some classifications of essential phenomena. The original classifications were really simple, even trivial, when presented. However, they suggested specific systematic relationships to be explored. This led, in later papers, to thorough empirical work and theoretical development. This is an example of how proper analytic models may aid research in a specific area, such as information seeking.

## **Discussion and conclusions**

The previous section presented a framework for information seeking studies that directly suggested research questions and hypotheses for testing. Such frameworks are clearly needed in building up a knowledge base in the IS&R domain. Unfortunately, the work discussed above is not complete and we cannot present a well thought-out complete framework. There is room for further work, which is not the purpose of the present paper. Moreover, the model discussed is very specific, it does not attempt to cover all phenomena related to [task-based] information seeking.

However, as a small contribution to further development, we can point to the fact that the model makes no reference to the characteristics of the person (apart from the possibility that novices and experts will behave differently), or to the field in which the person works. Other investigations have drawn attention to individual personality as a determinant of information-seeking behaviour (e.g., Kernan & Mojena, 1973; Bellardo, 1985; Palmer, 1991), and to the discipline or context within which the person works (e.g., Anon., 1965; Auster & Choo, 1994; Fabritius, 1998; Greene & Loughridge, 1996; Herner & Herner, 1967; Siatri, 1998; Timko & Loynes, 1989; Wilson & Streatfield, 1980). For example, the fact that more complex decisions involve more searching for people as sources of information may differ depending upon the person's 'need for affiliation' (McClelland, 1961).

From the point of view of context or discipline, even in the field of public administration, for example, there may be significant differences in the nature of the tasks in, say, a planning department and a more 'people oriented' department such as social work. In the former, the processing of applications may involve much more decision making of a formal, technical nature, while in the latter, the concern with people's personal and domestic problems may result in decisions that

have consequences that are more difficult to assess. We can suggest, therefore, a distinction between decisions that are related to a 'concern for process' and those that are related to a 'concern for person'.

We can also note that the distinction between 'information' and 'advice' is not sufficiently explored, although we suspect that the increased use of people as sources in complex decisions may have as much to do with the ability of people to guide, evaluate and advise, as with their possession of expert knowledge. Previous work on the affective dimension of information behaviour may also be relevant here (Wilson, 1981; Kuhlthau, 1993).

Finally, we can also point to a second dimension of decisions: as noted above, the present framework uses one dimension "*a priori* determinability of, or uncertainty about, task outcomes, process and information requirements". Thompson (1967) proposed two dimensions, one of which is similar to that used here, "Preference regarding possible outcomes", which might be 'certainty' or 'uncertainty'. The second dimension is "Beliefs about cause/effect relationships", which, again, might be 'certain' or 'uncertain'. The matrix that results from the combination of these two dimensions gives four types of decision processes, as shown in Figure 6.

Beliefs about cause/effect relations	Preferences regarding possible outcomes	
	Certainty	Uncertainty
Certain	<i>Computation</i>	<i>Compromise</i>
Uncertain	<i>Judgement</i>	<i>Inspiration</i>

**Figure 6: Decision processes (based on Thompson, 1967)**

The conceptual richness that results from the addition of a second dimension would give rise to an additional set of hypotheses relating the use of information sources to decision process. For example, one might hypothesise that decisions requiring 'judgement' will involve more information seeking activity and a greater use of discussions with colleagues, than other types of decision process, while 'inspiration' may require more personal 'thinking time' and use of a greater variety of information sources.

We return to the requirements on conceptual frameworks presented above. The framework developed by Byström, Järvelin and Vakkari, through several studies, may be claimed to meet several of the requirements. In Engelbart's terms, it suggests that tasks, information, and information channels and sources are central objects in information seeking. It further suggests how these objects are related to each other. The hypotheses generated were (are) fruitful goals for further research.

Regarding Bunge's (1967) functions for scientific theories, here applied for assessing conceptual frameworks, we find the following when assessing the Byström, Järvelin and Vakkari framework:

- Systematisation of knowledge by:
  - Integrating formerly separate parts of knowledge: *Task complexity studies from organisational research are integrated with information seeking studies.*
  - Generalising and explaining lower abstraction level knowledge (or observations, data) through higher level constructs: *Specific information needed and sought, much studied in information seeking, is analysed in terms of types of information.*
  - Explanation of facts through systems of hypotheses, which entail the facts: *The framework suggested and allowed verification of several hypotheses of the research domain, cf. Byström's S1-S11.*
  - Expanding knowledge by deducing new propositions based on selected starting points and collected information: *The later empirical and theoretical developments clearly expanded the original approach - Vakkari and Kuokkanen added to the original findings and unit theory.*
  - Improving the testability of hypotheses through the control context provided by systems of hypotheses: *The classifications generated many related hypotheses (e.g., Byström's S1 - S11) which provided, for each hypothesis, a context for its verification.*
- Guiding research by:
  - Pointing out fruitful problems: *From the beginning, the framework saw tasks, as opposed to whole jobs, related to information seeking through the types of information needed in tasks. The latter were seen to vary along task complexity.*
  - Proposing the collection of data, which nobody would think to collect without the theory: *The framework suggested data to be collected on task complexity, task-related information seeking and the types of information needed. These were novel ideas in the late 1980s in information seeking research.*
  - Proposing totally new lines of research: *The framework was one approach, among others, towards the task-centred line of information seeking research.*
- Mapping an area of reality by:
  - Representing or modelling the objects (and relationships) of that area instead of just summarising the data: *While early research in information seeking summarised job-level information seeking, sources and preferences, the framework suggested tasks and information types as explaining the phenomena..*
  - Providing a tool for producing new data: *The framework was a useful tool for generating hypotheses, and the associated research methods allowed the production of the required data.*

Regarding general scientific principles, suggested above, for the assessment of conceptual frameworks, we may point out the following:

- The framework is general in the sense that it supports the analysis of task-based information seeking for any kinds of tasks through categories that are not limited to special contexts, for example, academics. The tasks need not be job-related, leisure tasks do as well.
- The framework suggests perceived tasks, needed information and information seeking as a meaningful system. From the person or actor viewpoint this is much more meaningful than the information source and system framework (of many earlier studies) alone.

Further desiderata for conceptual models were:

- Simplicity: simpler is better other things being equal. *The framework is based on very simple classifications.*
- Accuracy: accuracy and explicitness in concepts is desirable. *The framework could be more accurate and explicit in its classification on task complexity. Nevertheless, it has functioned well as a first approximation. The framework is more accurate than its predecessors in its focus on task-level instead of job-level.*
- Scope: a broader scope is better because it subsumes narrower ones, other things being equal. *The framework is broader in its hospitality to any kind of tasks, not just job-related. On the other hand, it covers just three concepts, albeit important ones, of information seeking - a broader framework would incorporate other concepts as discussed above.*
- Systematic power: the ability to organise concepts, relationships and data in meaningful systematic ways is desirable. *This clearly is one strong feature of the framework.*
- Explanatory power: the ability to explain phenomena reliably and to predict them is desirable. *This clearly is one strong feature of the framework; it suggested several hypotheses that were later confirmed.*
- Validity: the ability to provide valid representations and findings is desirable. *(No model can directly argue for being valid)*
- Fruitfulness: the ability to suggest problems for solving and hypotheses for testing is desirable. *The number of studies that followed suggests at least some fruitfulness.*

We do not wish to make any claims about the usefulness or significance of this framework in comparison to other approaches within information seeking research. Rather, we wish to point out its formal merits: because of its characteristics, it has been successful in generating research that seems to have led to empirical and theoretical developments in the area of information seeking. Such models are needed in information science. According to Vakkari and Kuokkanen (1997), in order to create new knowledge in information science, we need clear, conceptually structured descriptions of the research objects. Without them the utilisation of

research results in further studies is hampered. That would lead to slow or non-existent growth of knowledge in the field while findings may still amass.

## Notes

1. It is this factor of determinability that helps us to define the 'automatic information processing' task. In such a task the outcome is determinable in advance. While a computer may be programmed to undertake tasks, which are computationally simple, such as those in a chess game, the outcome of the computer's calculations will not be determinable in advance, because of the complexity of the game.
2. Murtonen is the maiden name of Byström.

## References

- Anon. (1965). Survey of information needs of physicists and chemists. *Journal of Documentation*, **21**(2), 83-112.
- Anon. (1974). *Tietosysteemin rakentaminen* [Information system design]. Helsinki: Tietojenkäsittelyliitto. (Publication no. 25). (In Finnish).
- Auster, E., & Choo, C. W. (1994). How senior managers acquire and use information in environmental scanning. *Information Processing and Management*, **30**(5), 607-618.
- Barr, A. & Feigenbaum, E., (Eds.), (1981). *Handbook of artificial intelligence: Volume I*. London: Pitman.
- Belkin, N.J. (1980). Anomalous state of knowledge for information retrieval. *Canadian Journal of Information Science*, **5**, 133-143.
- Belkin, N.J. Cool, C., Stein, A., & Thiel, U. (1995). Cases, scripts and information seeking strategies: on the design of interactive information retrieval systems. *Expert Systems with Application*, **9**(3), 379-395.
- Bellardo, T. (1985). An investigation of online searcher traits and their relationship to search outcome. *Journal of the American Society for Information Science*, **36**(4), 241-250
- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, **56**(1), 71-90.
- Borlund, P. & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, **53**(3), 225-250.
- Borlund, P. & Ingwersen, P. (1998). Measures of relative relevance and ranked half-life: performance indicators for interactive IR. In: W.B. Croft, A. Moffat, C.J. van Rijsbergen, R. Wilkinson & J. Zobel (eds.), *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY: Association for Computing Machinery: 324-331.

- Bunge, M.A. (1967). *Scientific research*. (2 vols.) Heidelberg: Springer-Verlag.
- Byström, K. (1999). *Task complexity, information types and information sources*. Doctoral Dissertation. Tampere: University of Tampere. (Acta Universitatis Tamperensis 688).
- Byström, K. & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, **31**(2), 191 - 213.
- Campbell, D.J. (1988). Task complexity: a review and analysis. *Academy of Management Review*, **13**(1), 40-52.
- Daft, R.L. & Sormunen, J. & Parks, D. (1988). Chief executive scanning, environmental characteristics, and company performance: an empirical study. *Strategic Management Journal*, **9**(2), 123-139.
- Dervin, B. & Nilan, M. (1986). Information needs and uses. *Annual review of information science and technology*, **21**, 3-33.
- Ellis, D. (1989). A behavioural approach to information retrieval design. *Journal of Documentation*, **46**(3), 318-338.
- Ellis, D. & Cox, D. & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, **49**(4), 356-369.
- Engelbart, D.C. (1962). Augmenting human intellect: a conceptual framework. Menlo Park, CA: Stanford Research Institute. (Summary report AFOSR-3233) Retrieved 27 September 2003 from <http://www.bootstrap.org/augdocs/friedewald030402/augmentinghumanintellect/ahi62index.html>
- Fabritius, H. (1998). Information seeking in the newsroom. Application of the cognitive framework for analysis of the work context. *Information Research*, **4**(2). Retrieved 27 September, 2003, from <http://informationr.net/ir/4-2/isic/fabritiu.html>
- Fischer, W.A. (1979). The acquisition of technical information by R&D managers for problem solving in nonroutine contingency situations. *IEEE Transactions on Engineering Management*, **26**(1), 8-14.
- Fiske, D.W. & Maddi, S.R. (1961). Functions of varied-experience. Homewood, IL: Dorsey Press.
- Greene, F. & Loughbridge, B. (1996). Investigating the management information needs of academic Heads of Department: a Critical Success Factors approach. *Information Research*, **1**(3) Retrieved 27 September, 2003, from <http://informationr.net/ir/1-3/paper8.html>
- Hackman, J.R. (1969). Toward understanding the role of tasks in behavioral research. *Acta Psychologica*, **31**, 97-128.
- Hart, P.J. & Rice, R.E. (1991). Using information from external databases: contextual relationships of use, access method, task, database type, organizational differences, and outcomes. *Information Processing & Management*, **27**(5), 461-479.
- Herner, S., & Herner, M. (1967). Information needs and uses in science and technology. *Annual Review of Information Science and Technology*, **2**, 1-34.
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction. *Journal of Documentation*, **52**(1), 3-50.

- Järvelin, K. (1986). On information, information technology and the development of society: an information science perspective. In P. Ingwersen, L. Kajberg, & A. Mark Pejtersen (eds.), *Information technology and information use: towards a unified view of information and information technology*. London: Taylor Graham: 35-55.
- Järvelin, K. (1987). Kaksi yksinkertaista jäsennystä tiedon hankinnan tutkimusta varten [Two simple conceptual frameworks for information seeking research]. *Kirjastotiede ja Informatiikka*, **6**(1), 18-24. [In Finnish, English abstract]
- Järvelin, K. & Repo, A. (1983). On the impacts of modern information technology on information needs and seeking: a framework. In H.J. Dietschmann, (Ed.), *Representation and exchange of knowledge as a basis of information processes* (pp. 207-230). Amsterdam, NL: North-Holland.
- Järvelin, K. & Repo, A. (1984). A taxonomy of knowledge work support tools. *Proceedings of the Annual Meeting of the American Society for Information Science*, **21**, 59-62.
- Kernan, J.B., & Mojena, R. (1973). Information utilization and personality. *Journal of Communication*, **23**(3), 315-327
- Kuhlthau, C.C. (1991). Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, **42**(5): 361-371.
- Kuhlthau, C.C. (1993). *Seeking meaning: a process approach to library and information services*. Norwood, NY: Ablex.
- McClelland, D.C. (1961). *The achieving society*. New York, NY: Van Nostrand,
- MacMullin, S.E. & Taylor, R.S. (1984). Problem dimensions and information traits. *The Information Society* **3**(1), 91-111.
- March, J. & Simon, H. (1967). *Organizations*. (2nd ed.). New York, NY: Wiley.
- Murtonen, K. (1992). Tuloksellisempaan tiedonhankintatutkimukseen: prosessianalyysi tiedontarpeiden ja tiedonhankinnan tutkimuksessa [Toward more effective information seeking studies: use of process-analysis in information needs and information seeking research]. *Kirjastotiede ja Informatiikka* **11**(2), 43-52. (In Finnish)
- Murtonen, K. (1994). *Ammatilliset tiedontarpeet ja tiedonhankinta tutkimuskohteena: Tutkimus tehtävän kompleksisuuden vaikutuksista tiedontarpeisiin ja tiedonhankintaan*. [Professional information needs and information seeking as study objects: a study on the effects of task complexity on information needs and information seeking]. Thesis for the Degree of Licentiate of Social Sciences. Tampere: University of Tampere, Department of Information Studies. (In Finnish)
- Palmer, J. (1991). Scientists and information. II. Personal factors in information behaviour. *Journal of Documentation*, **47**(3), 254-275
- Roberts, K.H. & Glick, W. (1981). The job characteristics approach to task design: a critical review. *Journal of Applied Psychology*, **66**(2), 193-217.

- Saracevic, T. (1996). Modeling interaction in information retrieval: a review and proposal. *Proceedings of the Annual Academy Meeting of American Society for Information Science*, **33**, 3-9.
- Siatri, R. (1998). Information seeking in electronic environment: a comparative investigation among computer scientists in British and Greek universities. *Information Research*, **4**(2). Retrieved 27 September, 2003, from <http://informationr.net/ir/4-2/isic/siatri.html>
- Spink, A. (1997). Study of interactive feedback during mediated information retrieval. *Journal of the American Society for Information Science*, **48**(5), 382-394.
- Thompson, J.D. (1967). *Organizations in action*. New York, NY: McGraw-Hill Book Co.
- Tiamiyu, M.A. (1992). The relationships between source use and work complexity, decision-maker discretion and activity duration in Nigerian government ministries. *International Journal of Information Management*, **12**(2), 130-141.
- Timko, M., & Loynes, R.M.A. (1989). Market information needs for prairie farmers. *Canadian Journal of Agricultural Economics*, **37**, 609-627.
- Tushman, M.L. (1978). Technical communication in R&D laboratories: the impact of project work characteristics. *Academy of Management Journal*, **21**(4), 624-645.
- Vakkari, P. (1998). Growth of theories on information seeking. An analysis of growth of a theoretical research program on relation between task complexity and information seeking. *Information Processing & Management*, **34**(3/4), 361-382.
- Vakkari, P. (1999). Task complexity, problem structure and information actions. Integrating studies on information seeking and retrieval. *Information Processing & Management*, **35**(6), 819-837.
- 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*, **53**(5), 497-519.
- Van de Ven, A. & Ferry, D. (1980). *Measuring and assessing organizations*. New York, NY: Wiley.
- Wagner, D. & Berger, J. (1985). Do sociological theories grow? *American Journal of Sociology* **90**, 697-728.
- Wilson, T.D. (1981). On user studies and information needs. *Journal of Documentation*, **37**(1), 3-15.
- Wilson, T.D. (1997). Information behaviour: an interdisciplinary perspective. *Information Processing & Management*, **33**(4), 551-572.
- Wilson, T.D. (1999). Models in information behaviour research. *Journal of Documentation*, **55**(3), 249-270.
- Wilson, T.D. & Streatfield, D.R. (1980). "You can observe a lot..." A study of information use in local authority social services departments conducted by Project INISS Sheffield: University of Sheffield, Postgraduate School of Librarianship and Information Science. (Occasional Publication No. 12) Retrieved 27 September, 2003, from <http://informationr.net/tdw/publ/INISS/>

- Wood, R.E. (1986). Task complexity: definition of the construct. *Organizational Behavior and Human Decision Processes* 37(1), 60-82.
  - Zeffane, R.M. & Gul, F.A. (1993). The effects of task characteristics and sub-unit structure on dimensions of information processing. *Information Processing & Management*. 29(6): 703-719.
- 

## Appendix

Byström's (1999) eleven statements - cf. Figure 5.

S1: as soon as information acquisition requires an effort people as sources are more popular than documentary sources.

S2: the more information types are needed, the greater the share of people as sources.

S3: the more information types are needed, the greater the share of general-purpose sources and the smaller the share of task-oriented sources.

S4: the more information types are needed, the more sources are used.

S5: the internality of different source types is loosely connected to the information types.

S6: the higher the degree of task complexity, the more probable is the need for multiple information types: first task information, then task and domain information, and finally task, domain and [problem] solving information.

S7: the higher the degree of task complexity, the more information types are needed, and the greater the share of people as sources and the smaller the share of documentary sources.

S8: the higher the degree of task complexity, the more information types are needed and the greater the share of general-purpose sources and the smaller the share of task-oriented sources.

S9: the higher the degree of task complexity, the more information types are needed, and the higher the number of sources used.

S10: task complexity is distinctly related to increasing internality of people as sources and decreasing internality of documentary sources.

S11: Increasing task complexity fosters the use of people as sources.

[http://www.uh.cu/facultades/fcom/portal/interes\\_glosa\\_fuentes.htm](http://www.uh.cu/facultades/fcom/portal/interes_glosa_fuentes.htm) -->

## **Glosario de Tipología de Fuentes Documentales**

El mejor uso de la biblioteca se da cuando se pueden ofrecer servicios de referencia e información sobre el contenido de los fondos bibliográficos. También es útil conocer cuáles son las fuentes de información en las que debemos y/o podemos indagar. De acuerdo al tipo de información que contienen y la forma en que ésta se estructura podemos distinguir: fuentes de información: primarias, secundarias y terciarias.

**Fuentes de información primaria:** es aquella que genera, contiene, transfiere o suministra información original; resultante de un proceso intelectivo de investigación, creación o desarrollo.

Las fuentes de información primarias contienen información original. Su objetivo es comunicar los resultados del conocimiento y de la creación. Constituyen la colección básica de una biblioteca.

Pueden ser LIBROS, REVISTAS, TESIS DE DOCTORADO, LITERATURA GRIS, PATENTES DE INVENCION, etc. Pueden estar en distintos formatos, ya sea impreso, como libros, revistas, etc. o en formatos especiales, como videocasetes, discos compactos, etc.

### **LIBROS**

En cuanto a la forma, es un material impreso, no periódico, de muchas páginas. En cuanto a su contenido, es una pieza científica, profesional o literaria.

#### Características:

Contienen las generalizaciones principales de los estudios teóricos de los problemas cruciales de la ciencia. La información que contienen a menudo data de años atrás lo que hay que tener en cuenta según el área temática en la que estemos trabajando. Algunos tipos especiales de libros por su contenido son:

**TEXTOS Y MANUALES** están destinados principalmente a la enseñanza  
**CITAS DE CONGRESOS Y CONFERENCIAS** son los documentos que contienen los resúmenes y a veces las ponencias presentadas, así como las conclusiones de un congreso.

### **REVISTAS**

Publicación periódica que aparece regularmente, y puede tener distinta periodicidad (semanal, mensual, bimestral, etc.). Es el medio por excelencia donde se publican los avances científicos y las investigaciones en curso. La información

que contienen los artículos de revistas es incorporada a los libros mucho tiempo después de haber sido publicada la revista. Contienen menos generalizaciones y por ello pierden pronto actualidad. Tienen formato idéntico en todas sus ediciones.

### **PREPRINTS y OFFPRINTS**

Son los artículos de revistas que nos envía el propio autor cuando se le solicita copia. Los preprints son tiradas aparte de un artículo o un documento, impresas antes de sus publicaciones oficiales y destinadas a ser enviadas a un número limitado de personas interesadas. Los offprints son las ediciones separadas de un artículo que se entregan a los autores para enviarlas a sus colegas.

### **TESIS DE DOCTORADO**

Son trabajos de investigación originales que presentan los aspirantes a obtener un título de doctorado. En general es material inédito, por lo tanto se reproducen en pocos ejemplares (2 o 3).

### **LITERATURA GRIS**

Es aquella literatura de difícil acceso por no estar editada comercialmente. Por ejemplo las memorias, informes internos, publicaciones de gobierno.

**FUENTES DE INFORMACIÓN SECUNDARIAS:** es aquella que contiene suministra o transfiere información resultante de cualquier tipo de procesamiento de información primaria. Contienen información primaria reelaborada, sintetizada y reorganizada, o remiten a ella. Están especialmente diseñadas para facilitar y maximizar el acceso a las fuentes primarias.

Tipos de fuentes de información secundarias: Fuentes secundarias que reelaboran, sintetizan y reordenan la información de las fuentes primarias. Estos son los **DICCIONARIOS, ENCICLOPEDIAS, DICCIONARIOS BIOGRAFICOS**

Fuentes secundarias que remiten a otras fuentes primarias: **BIBLIOGRAFIAS, INDICES DE REVISTAS, BASES DE DATOS BIBLIOGRAFICAS, REVISTAS DE RESUMENES O ABSTRACTS, etc.**

**Revistas de resúmenes o abstracts:** Incluyen resúmenes de artículos aparecidos en distintas revistas de un área temática particular. Los especialistas las utilizan como medio de actualización para mantenerse al tanto de los progresos significativos en sus áreas de interés y también como una herramienta para la localización de artículos específicos aparecidos retrospectivamente.

**Características:** aparecen regularmente. Hoy en día son más utilizadas en formato CD-ROM, en línea o por Internet que en formato papel, ya que el acceso electrónico asegura una mayor cobertura temporal y rapidez en la búsqueda y actualización. Junto con cada resumen proveen los datos específicos para ubicar la revista donde se publicó un artículo y muy frecuentemente ofrecen el servicio de envío del artículo. Como mínimo permiten realizar búsquedas por autor, título, tema.

**FUENTES DE INFORMACIÓN TERCIARIAS:** Contienen información sobre las fuentes secundarias y remiten a ellas. Es un tipo de herramienta que utiliza más el bibliotecario. Entran en esta categoría, por ejemplo, las bibliografías de bibliografías

## **CARACTERISTICAS DE LA LITERATURA EN LAS DISTINTAS AREAS DEL CONOCIMIENTO**

Las características y el uso de la fuentes de información primarias varían de acuerdo con al área del conocimiento.

**CIENCIAS SOCIALES:** sociología, ciencia política, economía, antropología, educación, derecho, psicología, historia Mayor producción que en la literatura de ciencias exactas y de humanidades Los investigadores utilizan tanto libros como artículos de revistas (varía mucho de acuerdo a la carrera)

**CIENCIAS HUMANAS:** filosofía, artes, lingüística, literatura, teología Bajo nivel de obsolescencia, ya que el conocimiento no es progresivo. Preponderancia del uso del libro sobre el artículo de revista.

**CIENCIAS EXACTAS Y NATURALES:** Rápido crecimiento de la literatura Alto nivel de obsolescencia, ya que el conocimiento en esta área es progresivo, superando al conocimiento anterior. Preferencia por el uso de artículos de revistas más que por el de libros

**Fuente de información publicada:** es aquella fuente que después de ser sometida a un proceso editorial es publicada, responsabilizándose con su edición alguna organización dedicada a ese fin. **Fuente de información no publicada:** es aquella fuente documental que, habiendo sido creada y pudiendo ser utilizada, no ha sido sometida a un proceso editorial y, por lo tanto, tampoco publicada. **Fuente de información documental:** todo objeto que genere, contenga, suministre o transfiera información y sea, además, considerado documentos.

**Documentos:** Cualquier objeto material que registre o fije algún conocimiento y pueda ser incluido en una recopilación

- **Documentos Primarios:** objeto material que registre o fije algún conocimiento resultante de un proceso intelectivo de investigación, creación o desarrollo
- **Documentos Secundarios:** objeto material que registre o fije algún conocimiento información resultante de cualquier tipo de procesamiento de información primaria.
- **Documentos Terciarios:** objeto material que registre o fije algún conocimiento información resultante de cualquier tipo de procesamiento de información secundaria.

**Fuente de información no documental:** cualquier objeto o sujeto que genere, contenga, suministre o transfiera información, siempre que el objeto sea excluible de la categoría documento. En esta categoría se encuentran:

- **Fuentes de información institucionales:** son aquellas que estando representadas por organizaciones o grupos de trabajo (colecciones) institucionalmente identificados, generan, contienen, suministran o transfieren información
- **Fuentes de información personales:** son aquellos sujetos que en su actividad práctico-social generan, contienen, suministran o transfieren conocimiento/información
- **Fuentes de información impersonales (materiales):** son aquellas que en su estructura o composición orgánica generan, contienen, transfieren información utilizable en la naturaleza y en la práctica social

Sin embargo, la información no debe suscribirse sólo al fondo bibliográfico que posee la biblioteca en cuestión; ninguna biblioteca es autosuficiente para satisfacer todas las demandas del usuario. Por ello, en la actualidad se puede y se debe informar sobre lo que otras bibliotecas pueden ofrecer. Gracias a la aplicación del campo informático al sistema de gestión de bibliotecas, es posible el acceso fácil y rápido a catálogos de bibliotecas de cualquier lugar del mundo en forma "on line" o mediante la utilización de bases de datos en CD ROM (discos ópticos cuya lectura se realiza mediante un ordenador personal). El investigador podrá hacer tantas búsquedas como quiera en cualquiera de las bases de datos disponibles, sin límite de tiempo y acotando un campo de investigación.

Todos estos instrumentos deben ser manejados de manera eficaz por el bibliotecario. El ideal estará en enseñar a los usuarios a localizar lo que solicitan y a manejar las fuentes de información.

### **Materiales fundamentales: Las obras de referencia**

La biblioteca tiene que estar preparada para responder de manera rápida y segura a preguntas sobre hechos y personas de actualidad, nombres y conceptos de países, ciudades, animales, plantas y a preguntas sobre datos numéricos y estadísticos.

La biblioteca además de ser un centro proveedor de documentos, es también un centro de información. Para ello es necesario formar una colección de referencia adecuada.

La colección de referencia sirve tanto para los bibliotecarios, como para los lectores, puesto que aquellos la necesitan no sólo para su trabajo de comunicación con los usuarios, sino para la organización de la colección principalmente en los trabajos de catalogación,

Las obras de referencia están hechas para contener muchas informaciones autónomas, y para organizarlas de manera que sean de fácil acceso. De esta manera, concentran en una fuente de información noticias cuyo conocimiento exigiría la lectura de muchas páginas.

La colección de referencia debe comprender obras generales que sirven de introducción al resto de la colección bibliográfica, y su lugar está en las primeras estanterías en la Biblioteca de libre acceso, o en la sala de lectura de carácter más general, en las que no son de libre acceso. En todo caso se trata siempre de una colección de libre acceso tan abierta como los catálogos de la Biblioteca.

Las Obras de referencia especializadas, por limitarse a una parte de la realidad o campo de saber, tienen un doble destino: en las Bibliotecas de libre acceso las obras de consulta pueden ir al principio de la sección correspondiente. Pero en las Bibliotecas que no son de libre acceso, las obras de referencia especializada deben ser de libre acceso y pasar a constituir parte de la sección de referencia.

**Entre las Obras de Referencia encontraremos:**

### **Catálogos**

**Definición:** Los catálogos identifican y describen libros que forman una colección concreta, ofreciéndonos además datos para su localización. Catalogar supone describir una obra en sus partes esenciales para poder identificarla y poder recuperarla posteriormente.

Por tanto los catálogos son listas ordenadas sistemáticamente de una colección de materiales bibliográficos, que además nos darán la ubicación de los mismos, siendo su misión doble:

- identificar los documentos por los datos consignados en su descripción
- localizar su ubicación en la biblioteca. Normalmente por medio de una clave o signatura topográfica.

Este segundo aspecto será lo que lo diferenciará de las bibliografías, haciendo que el catálogo proporcione la información más completa de todo el fondo bibliográfico. Los catálogos tal y como los conocemos hoy en día, son propios del siglo XVIII, y será en el siglo XIX cuando se comience a mostrar interés en su

publicación como instrumento de información del contenido de unas bibliotecas, a otros. Los primeros en publicarlas serían las Bibliotecas Nacionales más importantes del momento.

Con la aplicación de la tecnología informática a las bibliotecas a mediados del siglo XX, se ve la posibilidad de reproducir los catálogos en distintos soportes y actualizarlos y difundirlos rápidamente.

**Entre las funciones de los catálogos están:**

- Indicar los nombres de personas o entidades con responsabilidad intelectual en la elaboración de una obra (autores, colaboradores, traductores, prologuistas, etc.)
- Indicar los títulos de las obras que posee la Biblioteca
- Indicar las obras que posee sobre una determinada materia, serie, colección.
- Indicar las características sobre el contenido y presentación de una obra ( si se compone de más volumen)
- Reunir todas las obras sobre un mismo autor
- Indicar la ubicación de los libros en los estantes

**Los catálogos se diferencian:**

**Por su uso:**

- **Internos o públicos**

**Por su elaboración:**

- **Alfabético de obras y autores:** las fichas bibliográficas están ordenadas alfabéticamente por autores (autor personal, entidad, colaboradores) o por la primera palabra del título de las obras anónimas
- **Alfabético de materias:** las fichas están ordenadas por la materia de la que trata la obra.
- **Topográfico:** de carácter interno ordenadas por situación en la estantería
- **Sistemático:** las fichas están ordenadas por una notación determinada de conceptos.

**Tipos:**

- **Catálogos comerciales:** Recogen las publicaciones de los editores o impresores.
- **Individuales:** las publicaciones de cada editorial
- **Catálogos de bibliotecas:** recogen los fondos existentes en las distintas bibliotecas.

- **Colectivos:** entre estos destacan el NUC(National Union Catalogue); General Catalogue of Printed Books; Catalogue Générales des Livres Imprimés; Primmo Catalogo Colletivo delle Bibliothek Italiane; REBIUN(Catálogo colectivo de bibliotecas universitarias españolas); RUEDO (Unión de catálogos de 15 bibliotecas españolas con el mismo sistema de gestión bibliotecaria-DOBIS LIBIS. Consulta ONLINE exclusivamente); CCPB (Catálogo colectivo del patrimonio bibliográfico)
- Nacionales
- Regionales
- Locales
- Especializados
- De grandes bibliotecas: normalmente suelen ser las nacionales (en muchos casos se puede acceder a esa información en forma on-line o Internet)
- De bibliotecas individuales
- Catálogos de publicaciones periódicas: facilitan información sobre el título de las publicaciones periódicas que se encuentran en una unidad de información. Habitualmente, se completa con la relación de los años cuyos fascículos se encuentran en esa unidad.

### **Bibliografías**

**Definición:** las bibliografías buscan, identifican, describen conjuntos de libros que no forman una colección determinada y cuyas noticias se presentan debidamente ordenadas por medio de algunos de los elementos de la noticia.

#### **Tipos:**

**Bibliografías de bibliografías:** son la suma o recopilación de repertorios bibliográficos. Constituyen el primer paso en la búsqueda bibliográfica para una investigación especializada. Su función es servir de guía para señalar la existencia de bibliografías sobre la materia de la que queremos realizar el estudio.

Existen tantas clases de bibliografías de bibliografías como bibliografías hay:

- Universales
- Especiales
- Bibliografías universales:
- En curso
- Acumulativas
- Retrospectivas
- Bibliografías nacionales generales: el concepto no es idéntico en todos los países, ya que aunque persiguen el mismo objetivo obedecen a particularismos nacionales que han fijado la personalidad de estas bibliografías.

Hay que hacer hincapié en tres criterios: lengua, país de origen y naturaleza del documento.

Podemos decir que son las Bibliotecas Nacionales, de carácter general, las encargadas de elaborarlas en virtud de disposiciones legales. Repertoriando todo lo publicado en el territorio de una nación sea cual sea su soporte, tema o lengua del país. Además de lo publicado en el extranjero sobre el país o por autores del país.

Habitualmente se publican en Boletines mensuales; además de las publicaciones de los circuitos comerciales, recoge también lo de fuera de ellos (Boletines oficiales, Boletines de Instituciones y la denominada Literatura Gris).

**Entre ellas se encuentran:** La Bibliografía Nacional Española, la Deutsche National Bibliographie, la British National Bibliography, la Bibliographie Nationale Française y la Bibliografia Nazionale Italiana.

Las tres primeras bases de datos tienen el mismo programa de recuperación de la información, lo que facilita el acceso a la misma y la cooperación interbibliotecaria.

#### **Tipos:**

- **En curso o periódica:** son puntuales, con periodicidad fija señalando las nuevas incorporaciones y su finalidad es la de informar sobre la actualidad de lo que se publica en un país.
- **Acumulativas**
- **Retrospectivas:** en general son selectivas, orientadas a la investigación de tipo histórico. Son inventarios de la producción impresa de épocas pasadas en un país determinado.
- **Bibliografías Retrospectivas:** Se encargan de recoger obras que están fuera del comercio, tanto de editoriales como de particulares. Normalmente los elaboran los libreros anticuarios.
- **Entre los más conocidos encontraremos:** Book Auction Record, Cuadernos de Bibliofilia, L'Argus du Livre Ancienne et Moderne.
- **Bibliografías Comerciales. Catálogos de libreros:** son repertorios por y/o para el comercio del libro en un país. Elaborados por asociaciones profesionales con la adopción de normas internacionales de elaboración (ISBD-ISBN, ISSN).

Se trata de catálogos de editores con información sobre los libros que se hallan a la venta en el momento de su publicación.

Podemos decir que los catálogos comerciales existían ya en Grecia y Roma, donde los libreros daban a conocer sus existencias mediante una especie de inventario de los libros que poseían conocidos como BIBLIOPOLA.

En la Edad Media con la invención de la imprenta el comercio de los libros comienza alcanzar un volumen considerable. Los libreros comenzaran a editar sus catálogos (MESSKATALOGUE), con el fin de vender sus existencias. Al principio,

sólo eran listas de libros, pero con la aparición de obras con mayor información tienden a convertirse en bibliografías detalladas.

En el siglo XVII, se dará una consolidación de este tipo de catálogos; en el XVIII, se sistematizará, y en el siglo XIX, los editores darán a conocer sus existencias con periodicidad.

A finales del siglo XIX, tenemos dos bibliografías comerciales de mucho interés en EEUU: Publishers weekly y Cummulative Book Index. En el siglo XX, se crearán asociaciones de libreros y distribuidores para elaborar catálogos colectivos, debido a la necesidad de la puesta al día constante.

Al margen de EEUU, en el ámbito europeo también tendremos repertorios comerciales editados con periodicidad. Entre estos:

- Libros Españoles en Venta; su formato CD-ROM se denomina CD-LID
- Livres Disponibles; su formato CD-ROM es ELECTRE
- Deutsche Verzeichnis
- British Books in Print

**Bibliografías de Publicaciones Periódicas:** son repertorios que recogen títulos de publicaciones periódicas entendidas como: "publicaciones impresas o no, que aparecen en volúmenes o fascículos sucesivos, ligados por lo general numérica o cronológicamente durante un tiempo no limitado de antemano".

**Bibliografías de Publicaciones Oficiales y Académicas:** se trata de repertorios que recogen escritos y documentos gubernamentales editados por orden, y a expensas de cualquier autoridad pública, como: decretos, leyes, tratados, memorandums, boletines oficiales, etc.

**Bibliografías especiales o especializadas:** contienen información especializada por:

- Por la forma o tipo del documento
- Por la materia o cobertura temática
- Por el tiempo
- Topobibliografías

**Las Bibliografías especializadas abarcan:** Monografías, Publicaciones periódicas, Patentes, Comunicados oficiales...

Existen tanto en soporte impreso como en CD ROM y suelen hacer vacíos de artículos de revistas que traten sobre el tema de su especialidad.

**Entre las más conocidas e importantes estarán:** MLA (Modern Language Association) especializada en Lengua y Literatura Moderna, contiene básicamente

artículos de publicaciones periódicas. SOCIOFILE, especializada en el área de Sociología. DISSERTATION ABSTRACT ON DISC, que recogen exclusivamente tesis doctorales.

### **Diccionarios**

**Definición:** Listas de palabras ordenadas alfabéticamente y con su significación. Un diccionario puede adoptar muchas formas (etimológico, histórico, bilingüe, políglota, ...) en función del mayor o menor detalle en la explicación de las palabras, en su historia, en la forma de utilización de las mismas, en las muestras de uso, así como la presencia o no de formas equivalentes en otros idiomas.

#### **Tipos:**

**Diccionarios enciclopédicos:** Son diccionarios en los que a la definición del significado de la palabra se une la de la cosa significada. Es diccionario por definir palabras, y enciclopédico porque la descripción se extiende a todos los conocimientos. Su universalidad y el hecho de que su ordenación se base sobre el idioma, hacen de esta forma documental, la principal entre las obras de referencia.

Son imprescindibles, el de la Real Academia Española y el ideológico de Casares, así como los más conocidos de las diversas lenguas. También los especializados son de suma utilidad, como el de legislación de Aranzadi, el geográfico de Villalba, el de Heráldica vasca de Querexeta, y el diccionario de euskera en nuestro caso.

**Diccionarios biográficos:** Aunque los nombres propios forman parte de las palabras contenidas en los diccionarios enciclopédicos, los diccionarios biográficos son esenciales para la solución de preguntas por parte de los lectores. Los diccionarios biográficos contienen resúmenes biográficos, ordenados alfabéticamente por el nombre del biografiado. Hay mucha variedad de ellos, ya que pueden ser nacionales o internacionales, generales o especializados, en curso o retrospectivos. También son obras de suma importancia y que pueden servir de arranque a trabajos de investigación.

La dificultad en España está en la dispersión de este tipo de obras, donde además falta el Gran diccionario biográfico retrospectivo.

### **Directorios**

**Definición:** Listas de personas u organismos con las informaciones oportunas sobre los mismos. Son imprescindibles para mantener la colaboración con otros centros.

### **Enciclopedias**

**Definición:** Prescinden de los aspectos lingüísticos y contienen el conjunto de todos los conocimientos humanos ordenados alfabéticamente, o temáticamente.

La enorme movilidad del conocimiento humano, han restado importancia a este tipo de obras, de igual manera que a un conjunto de manuales o tratados. Esto se ve claramente en otro tipo de obras cercanas a las colecciones enciclopédicas que dotadas de un tomo de índices resultan de suma utilidad.

Las encyclopedias constituyen un recurso al que acude con frecuencia la industria editorial española. El problema es pues, el de la selección. Esta debe hacerse en base a características como:

- Formato acertado
- Calidad gráfica del texto e ilustraciones
- Normalización en la grafía
- Cantidad y calidad de información, que estará garantizada por la autoridad de sus colaboradores, la actualidad y el equilibrio entre las materias.

Entre las más conocidas figuran:

- La Encyclopedie Universal Ilustrada de Espasa Calpe: la más completa en lengua española.
- La Gran Encyclopedie Rialp.
- El Diccionario Encyclopédico Larousse, de la editorial Planeta, ordenado alfabéticamente.
- La Encyclopedie Británica, de tipo científico con ordenación temática y buenos índices.
- La Encyclopedie Italiana, imprescindible en el campo humanístico.

Dentro del País vasco, contamos con el Diccionario Encyclopédico Ilustrado del País vasco de Auñamendi.

### **Manuales, Tratados y Atlas**

Constituyen una buena fuente de información, su único problema es que deben mantenerse al día.

### **Anuarios, Memorias, Estadísticas**

**Definición:** Son el resultado de la gestión de los organismos oficiales generalmente, como los del Instituto Nacional de Estadística, los Anuarios del Gobierno vasco, Memorias de las Diputaciones, de las entidades de ahorro, Universidades, etc. En su mayoría son publicaciones de periodicidad anual.

Es importante decir que en este documento solo aparece alguna de las clasificaciones de fuentes de información, pero no son las únicas. Si desea abundar sobre el tema revise la bibliografía utilizada para la confección de este trabajo que a continuación se expone:

## **Bibliografía:**

Cruz-Paz, Andrés Fuentes de información: Aspectos teóricos. -- La Habana: Universidad de la Habana, 1994. -- 104 p.

Félix del Valle Gastaminza Documento. Concepto y tipología  
<<http://www.ucm.es/info/multidoc/prof/fvalle/tema3.htm>> [Consulta: 27 junio 2001]

Uria Fernández, Irune. Búsqueda en la biblioteca / Irune Uria Fernández, Sol María Menéndez Otamendi, Maite Agirreamalloa Alzola< <http://www.serv-inf.deusto.es/abaitua/konzeptu/htxt/grupoe.htm> >

**[Consulta: 20 de junio de 2002]**

## Capítulo 3. Consultas de Bases de Datos

### Bases de Datos: conceptos, características, aplicaciones

La información es uno de los recursos más valorados en la sociedad actual, pero en los momentos actuales nos enfrentamos a los volúmenes de información, cuyo crecimiento es exponencial a raíz del surgimiento y rápida evolución de las tecnologías de la información y la comunicación y de la Internet. Este último pone al alcance de todos, un gran acervo de información de cualquier tipo. Sin embargo, la mayoría de las personas que buscan información en Internet pierden mucho tiempo en lograr su objetivo y finalmente obtienen resultados decepcionantes. De modo que, hoy en día es tan importante "tener acceso a la información relevante", como ser capaz de "descartar la irrelevante".

La necesidad de encontrar información pertinente, precisa y en el momento oportuno, en un gran volumen de documentos, que además son heterogéneos, provoca que los sistemas que permiten la recuperación de información estén en un momento crucial de cambio. Por otra parte, la necesidad de acercar la herramienta al usuario desencadena una amplia actividad investigadora acerca de las interfaces de usuario. Los tecnólogos y los especialistas en información deben estar preparados para sacar provecho de estos cambios y proponer los que aún han de mejorar los sistemas actuales, adelantándose a las expectativas de sus usuarios.

#### Bases de datos

En la actualidad, los sistemas de bases de datos se han convertido en el fundamento que garantiza el funcionamiento correcto de los sistemas de información para la gestión corporativa. La importancia de la información, su almacenamiento, manipulación y recuperación en forma eficiente, es vital y estratégico para cualquier organización. Este hecho ha propiciado el desenvolvimiento de los conceptos y de las técnicas para el manejo eficaz de los datos. Las bases de datos juegan un rol crítico en casi todas las áreas donde las computadoras son usadas, incluyendo negocios, ingeniería, medicina, leyes, educación, etc.

Además, la implantación de los cambios tecnológicos ha respondido a las exigencias del ámbito competitivo de las organizaciones modernas; de modo que se puede afirmar que existe un vínculo estrecho entre la tecnología de las bases de datos y las necesidades del actual mundo empresarial. Es necesario destacar que teniendo en cuenta el desarrollo desigual de aplicación de los sistemas computacionales en diferentes organizaciones, en la práctica moderna aún están presentes disímiles estadios en el desenvolvimiento, implantación y explotación de los sistemas de información.

Las bases de datos almacenan, como su nombre dice, datos. Estos datos son representaciones de sucesos y objetos, a diferente nivel, existentes en el mundo real: en su conjunto, representan algún tipo de entidad existente. En el mundo real se tiene percepción sobre las entidades u objetos y sobre los atributos de esos objetos; en el mundo de los datos, hay registros de eventos y datos de eventos. Además, en ambos escenarios se puede incluso distinguir una tercera faceta: aquella que comprende las definiciones de las entidades externas, o bien las definiciones de los registros y de los datos.

Los sistemas de bases de datos representan una disciplina compleja y en evolución, cuyo fundamento esencial se encuentra en los aspectos más generales del desarrollo y el uso de las bases de datos. La complejidad y el tamaño de las bases de datos varían grandemente, desde las bases de datos personales, diseñadas y utilizadas por una misma persona sobre una sola computadora, hasta las bases de datos de organizaciones, que modelan el flujo de información crítica de grandes empresas o instituciones, incluso, internacionales.

En cualquier sistema de información basado en computadoras se cuenta con dos enfoques diferentes para el almacenamiento y procesamiento de los datos.

1. Enfoque orientado a ficheros o archivos
2. Enfoque orientado a bases de datos

### **Enfoque orientado a ficheros o archivos**

Un Fichero de datos es una colección de datos estructurada de una forma particular y empleada con un determinado propósito, donde se almacena información sobre una cosa, un objeto o una entidad

Los ficheros o archivos de datos son una manera práctica de almacenar los datos necesarios para una aplicación informática. Estos se pueden diseñar y elaborar de forma rápida. Están formados por campos o datos que tienen un formato determinado y pos artículos o registros que contienen los datos.

Los ficheros se pueden organizar utilizando diferentes técnicas de acuerdo con el tipo de procesamiento que se desee realizar, siendo las formas de organización más comunes las siguientes:

- Secuencial: los artículos o registros son procesados secuencialmente (uno detrás de otro) en el mismo orden en que se encuentran almacenados en el fichero.
- Directa: los artículos o registros son procesados no en el orden en que se encuentran en el fichero sino de acuerdo a la posición que ocupan, utilizando alguna técnica de generación de direcciones.

- Indexada: los artículos o registros son organizados en un índice que contiene la dirección del registro y el valor del campo por el cual se indexó.

## Enfoque orientado a bases de datos

### Concepto de base de datos

En la literatura existen diferentes definiciones de base de datos, pero una de las más compactas es la brindada por Allen Taylor, quien plantea: “... *una base de datos es una colección auto-descriptiva de registros integrados*”.

**Auto-descriptiva**, puesto que la base de datos contiene los metadatos - la descripción de su estructura - en el catálogo o diccionario de datos de la propia base de datos; e **integrada**, porque no solo contiene los datos sino también las interrelaciones entre ellos. Y precisamente en estos dos elementos radica el valor de las bases de datos, que no está en los datos propiamente dichos sino en la estructura que modela la realidad, de modo que una base de datos puede procesarse por uno o más sistemas de aplicación.

Según Date, un “**sistema de base de datos (SBD)** no es más que un sistema computacional de mantenimiento de registros (computerized record-keeping system) y se diseña para manejar grandes cantidades de información”. El tratamiento de los datos conlleva tanto la definición de las estructuras para el almacenamiento de la información así como el suministro de mecanismos para la manipulación de la información.

Para Adoración de Miguel y Mario Piattini, “una base de datos es una colección de datos estructurados según un modelo que refleje las relaciones y restricciones existentes en el mundo real. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de éstas, y su definición y descripción han de ser únicas estando almacenadas junto a los mismos. Por último, los tratamientos que sufren estos datos tendrán que conservar la integridad y seguridad de éstos”.

El sistema de bases de datos brinda al usuario un conjunto de facilidades que permiten realizar diferentes operaciones sobre tales ficheros, entre que se encuentran:

- Añadir nuevos ficheros, vacíos, a la base de datos
- Insertar nuevos datos en los ficheros existentes
- Recuperar (consultar, obtener) datos de los ficheros existentes
- Modificar los datos en los ficheros existentes
- Eliminar datos de los ficheros existentes
- Suprimir ficheros existentes, vacíos o no, de la base de datos

En resumen, una BD es un almacenamiento de datos relacionados entre si, almacenados permanentemente y controlado centralmente para servir a múltiples y diferentes aplicaciones. No son una mera colección de datos o de archivos, sino una fuente central de datos significativos, los cuales son compartidos por varios usuarios para satisfacer diferentes necesidades. El compartir los datos significa que estos se almacenan solo una vez, manteniendo la integridad de los datos.

La esencia de una base de datos radica en el Sistema de Gestión de la BD (DBMS Database Management System) el cual permite la creación, modificación y recuperación de los datos existiendo una persona que es el administrador de la DB.

El enfoque de las BD tiene la ventaja principal de permitir que los usuarios expongan sus puntos de vista sobre los datos, sin preocuparse de la estructura y composición de los datos o de su ubicación física. Una BD bien diseñada es mucho más flexible que varios ficheros independientes.

Entre las fuentes de información referentes a bases de datos bibliográficas, como son Medline, Lisa, Excerpta Medica, Lilacs, entre otras, el sistema de búsqueda le permitirá al usuario seleccionar la fuente de información conforme a su decisión sobre la naturaleza del problema y el repositorio que el considere más adecuado a sus propósitos.

**Bases de datos en línea.** En la medida que la construcción de bases de datos textuales ha evolucionado conjuntamente con el relativo abaratamiento de las memorias masivas, las grandes bases de datos en línea han sido desarrolladas de modo que colectivamente contienen muchos millones de documentos de varios tipos, lo que representa un vasto almacén de información, que se puede consultar cada vez con mayor amplitud, comodidades y precisión.

Como ya se ha dicho el mundo de hoy se está aproximando, si ya no lo ha excedido, al punto de sobrecarga o exceso de información. Esto no quiere decir que no se utiliza de forma efectiva dicha información, sino que ese uso efectivo actual es solamente una pequeña fracción de la potencialidad que pueden brindar algunos mecanismos más apropiados de almacenamiento y recuperación.

Por su parte, la tendencia actual de los sistemas de bases de datos no se restringe a las aplicaciones tradicionales, tales como el control de inventarios, sino que se extiende a aplicaciones que se relacionan con el diseño automático, las bases de conocimientos, los sistemas de descubrimiento automático, el tratamiento de textos, imágenes, sonidos y otras, razón por la cual en el presente contexto se ha hecho referencia a **información** y no únicamente a datos. Esto implica que los nuevos usuarios de bases de datos también deberán contemplar una formación en estas tecnologías debido esencialmente a que todas las versiones del futuro se fundamentarán de una u otra manera en estos enfoques.

**FACULTAD DE TRADUCCIÓN Y DOCUMENTACIÓN  
UNIVERSIDAD DE SALAMANCA**

José Antonio Merlo Vega

**Fuentes de información electrónicas para la investigación en  
Biblioteconomía**

**Artículo publicado en la Revista Educación y Biblioteca, octubre 1997, n. 83,  
p. 48-56.**

## **0. Introducción**

En los últimos tiempos los investigadores de cualquier disciplina se han visto obligados a modificar sus hábitos en cuanto a la búsqueda y recuperación de información. Los productos y técnicas tradicionales, a pesar de seguir siendo válidos, están siendo relegados a un segundo plano debido a la eficacia, rapidez, exhaustividad y facilidad de uso de las fuentes de información automatizadas. Nadie duda hoy de las ventajas de las bases de datos frente a los grandes repertorios impresos, ni de la utilidad que Internet presenta para la investigación o la localización de documentos.

Distintos son los documentos en formato electrónico que existen en la actualidad. Explicar la gran cantidad de variantes y versiones que presentan estos soportes implicaría dedicar excesivo espacio y derivar la cuestión principal de este trabajo. Es preferible simplificar la tipología de los documentos electrónicos y citar únicamente las grandes familias de estos soportes. Coinciendo con Ernest Abadal los servicios de información electrónica pueden resumirse en: bases de datos en línea, videotexto, teletexto, discos ópticos y Web. De entre todos ellos destacan, por su demostrada utilidad, los CD-ROM y los documentos Web.

Este artículo se centrará la exposición y comentario de fuentes de información de interés para los profesionales de las bibliotecas que estén en los dos soportes anteriores. Los CD-ROM han sido elegidos por la expansión y el desarrollo que han tenido en los últimos años. En estos momentos, la mayor parte de las bases de datos en línea sobre Biblioteconomía se encuentran en disco óptico, por lo que era preferible citar únicamente los CD-ROM, evitando la duplicidad de información. Lógicamente, y por otra parte, se incluye también aquí la información World Wide Web, ya que Internet es en estos momentos una fuente de información de incalculable valor, puesto que reúne millones de documentos.

## **1. Productos en CD-ROM de interés para Biblioteconomía, Documentación y Bibliografía**

El CD-ROM es producto de gran utilidad para quienes por razones profesionales, académicas o de cualquier otro tipo se dedican a la investigación. Su gran

capacidad de almacenamiento y la rapidez en la recuperación de la información, hacen de éste una herramienta de elevada utilidad para la localización de documentos. Se encuentran en este soporte multitud de bases de datos bibliográficas o documentales de interés para la Biblioteconomía. En la selección que a continuación se acompaña se han agrupado los títulos dependiendo del contenido de los mismos, apareciendo en primer lugar las bases de datos sobre documentación existentes en CD-ROM, que son las mismas que se encuentran en línea, para después incluir otras fuentes de información bibliográfica de indudable interés, como son los catálogos de distintas bibliotecas o los repertorios de obras en venta.

### **1.1. Bases de datos de Biblioteconomía y Documentación**

Information Science Abstracts. Cobre todos los aspectos de la Biblioteconomía y las tecnologías de la información. Recoge más de 450 revistas internacionales, libros, patentes, actas de congresos e informes. A diferencia de Lisa cubre en profundidad los aspectos tecnológicos: telecomunicaciones, CD-ROM, online, robótica, redes hipermedia, etc. Recoge información de los últimos 25 años. Editada por SilverPlatter.

Library Literature. Editada por H.W.Wilson y Silver Platter. Indiza 250 publicaciones del sector, tanto de EE.UU. como extranjeras, así como monografías, actas de congresos, tesis, etc. desde 1984. Su base de datos cuenta con más de 100.000 referencias. Accesible vía Internet previa suscripción. <<http://www.hwwilson.com/liblit.html>>

LISA Plus. Editada por Bowker Saur, es la más importante base de datos en el mundo de la profesión. LISA resume más de 500 publicaciones periódicas de 68 países, más actas de congresos, tesis e informes desde 1969. También se incluye en este disco la base de datos CRLIS, Current Research in Library and Information Science, que da cuenta de los proyectos de investigación que se están llevando a cabo, a nivel mundial en el campo de la Biblioteconomía y la Documentación, desde 1981.

Pascal.Sciences de l'information. Pascal es una base de datos pluridisciplinar, producida por el INIST francés, que cuenta con 9 millones de referencias, procedentes de 20.000 publicaciones periódicas internacionales, desde 1987 hasta hoy. Uno de los ficheros de esta base esta dedicado a las Ciencias de la Información, analizando revistas, libros, congresos, etc.

ISOC-DC (CSIC). Es una de las bases de datos que realiza el CSIC. ISOC-DC, recoge los artículos de revista publicados en España desde 1975 hasta la actualidad sobre Biblioteconomía, Documentación, Bibliografía, Archivística y demás ciencias relacionadas con la información. Asimismo recoge las actas de Congresos sobre la profesión publicadas en nuestro país.

## **1.2. Bases de datos multidisciplinares con información sobre Documentación**

Dissertation abstracts. Editada por la University of Microfilm International, recoge cerca de 1.500.000 de citas de tesis leídas o publicadas en más de 1.000 universidades de todo el mundo, desde 1861. Incluye todos los campos del conocimiento.

DOC-Theses. Recoge más de 300.000 tesis doctorales francesas de todas las áreas científicas, desde 1972. Las específicas de salud, desde 1983. La distribuye Chadwyck Healey

SIGLE. Esta producida por EAGLE, European Association for Grey Literature Exploitation. Agrupa más de 365.000 registros de literatura gris, es decir, documentos que no están disponibles a través de canales comerciales. Abarca desde 1980 hasta la actualidad. Es distribuida por SilverPlatter.

## **1.3. Obras de consulta profesionales**

American Library Directory. 36.000 bibliotecas de todo tipo y de todas las materias del ámbito norteamericano, canadiense y mexicano. Editado por Silver Platter.

Autoridades de la Biblioteca Nacional de España. Cerca de 300.000 registros de autores, materias, títulos uniformes de la lista de autoridades de la Biblioteca Nacional. Editado por Chadwyck Healey.

Bibliographic Index. Bibliografía de bibliografías sobre todos los campos del conocimiento, divididas de acuerdo a los encabezamientos de materia de la Library of Congress. Editado por Wilson.

Boston Spa Conferences y Boston Spa Serials. The British Library Document Supply Center, recopila en el primero de estos CD-ROM más de 350.000 actas de congresos celebradas desde mediados del siglo XVIII en cualquier país y sobre cualquier ámbito. Los documentos están accesibles en este mismo centro. El segundo disco reúne más de 500.000 referencias de publicaciones periódicas recibidas en el BLDSC.

CD-CATTS. Sistema de catalogación en curso para bibliotecas, que agrupa más de millón y medio de registros de la Biblioteca del Congreso, la Biblioteca Nacional de Canadá, la Biblioteca Británica, la Biblioteca Nacional de Medicina de EE.UU. y los US Official Printing Office. Se realiza desde, desde 1984 y es distribuida por Utlas.

CDMARC Names, Serials y Subjects. Estos CD-ROM se han venido editando hasta 1996. Contienen, respectivamente, el fichero de autoridades de la Biblioteca del Congreso, con cerca de 3.500.000 de entradas, 700.000 registros de

publicaciones periódicas, de los fondos de la Biblioteca del Congreso y, en tercer lugar, el fichero de materias completo de la Biblioteca del Congreso de Washington.

CD ROM of CD ROMS. Enciclopedia multimedia sobre CD-ROMs y lo relativo a su industria. Más de 4.000 títulos y cerca de 5.000 empresas. Editado por Walnut Creek.

CD ROMS In print.- Editado por Gale Research, da información sobre 6.500 CD-ROMs, sus editores, distribuidores, software, etc.

Gale directory of databases. Como su nombre indica es un directorio de bases de datos, además de sobre la industria de la información electrónica. Engloba 6.000 bases disponibles a través de 800 servicios online internacionales, 3.000 bases en CD-ROM, disquete, cinta magnética, etc., 5.000 organizaciones, sistemas y servicios relacionados con la producción y distribución y 2.300 organizaciones de sistemas y servicios de telecomunicación. Silver Platter es su distribuidora.

ISSN.- Publicada por el Centro nacional del ISDS y Chadwyck, contiene registros de más de 650.000 publicaciones periódicas de 193 países y 144 lenguas. Así mismo recoge la List of Serials Title Word Abbreviations. La distribuye Chadwyck-Healey.

Multimedia and CD-ROM directory. Información sobre 6.000 títulos en CD-ROM y cerca de 4.000 empresas. Además da datos sobre hardware, software, etc. Editado por TPFL.

Myriade. Catálogo colectivo de publicaciones en serie francés. Recoge las publicaciones periódicas de 2.800 bibliotecas y centros de documentación. Contiene 250.000 registros en formato ISDS-MARC que permite su exportación . Lo edita Chadwyck Healey.

Les Notices d'Autorité de BN-OPALE. 550.000 registros de autores, títulos uniformes y materias de la Biblioteca Nacional de Francia, extraídos de los libros recibidos desde 1970. Editado por Chadwyck-Healey.

Periodical Context Index. Recoge la información de 15 millones de artículos de 3.500 revistas, editadas hasta 1990, de humanidades y ciencias sociales, divididas en 37 materias y con el sumario de cada fascículo. Editada por Chadwyck-Healey.

Publishing market place reference plus. Más de 83.000 bibliotecas mundiales, 32.000 librerías y 80.000 editores reunidos en esta base de datos editada por Bowker.

REPERE. 250.000 citas de artículos de 500 revistas publicadas en lengua francesa en Francia, Suiza, Canadá y Bélgica. Editado por Silver Platter.

Serials. Editada por Utlas, contiene información bibliográfica en formato MARC de más de 460.000 publicaciones seriadas disponibles desde antes de 1987. La información proviene de la Biblioteca del Congreso de los Estados Unidos y la Biblioteca Nacional de Canadá.

Serials directory.- 160.000 títulos de publicaciones periódicas y congresos. Presenta también datos de cerca de 87.000 editores. Los registros se pueden descargar en formato MARC e ISSN. Está editado por Ebsco.

Spanish Union Catalogue of Periodicals. Catálogo editado por Chadwyck Healey, que recoge publicaciones periódicas en bibliotecas españolas. La información se puede recuperar por título, ISSN, editor, etc.

Ulrich's Plus. Editada por Bowker Saur, recoge 200.000 publicaciones periódicas regulares y 50.000 irregulares. Además contiene información de 80.000 editores en 200 países, con 22 criterios de búsqueda distintos.

#### **1.4. Directorios de profesionales e instituciones**

Directorio de bibliotecas españolas. Relación de 9. 500 bibliotecas españolas. Se ofrece la dirección, especialización, tipo, fondos, etc. Lo distribuye Chadwyck Healey.

Directory of library and information professionals. Proporciona datos sobre cerca de 50.000 profesionales de la información y la documentación. Aparecen datos sobre su formación, experiencia profesional, publicaciones, áreas de especialización, etc. Esta editado por la American Library Association.

European book world. En tres CD-ROM se da información de 18.000 editores europeos y 7.500 organismos relacionados, 45.000 bibliotecas y 35.000 librerías del continente europeo. Lo edita Anderson Rand.

Library Reference Plus. Editado por Bowker, incluye información referencial completa sobre todas las bibliotecas norteamericanas, librerías, distribuidores, editores, mayoristas, comercio editorial, etc. Incluye el Directorio de Bibliotecas Americanas y el Directorio Americano de Comercio Editorial de Estados Unidos

World guide to libraries plus. Editada por Bowker Saur, da información de 47.000 instituciones en 18 países. Se incluyen todo tipo de bibliotecas con su dirección postal y teléfono.

#### **1.5. Bibliografías nacionales**

Bibliografía nacional española. Contiene los registros bibliográficos de los títulos recibidos por depósito legal en la Biblioteca Nacional desde 1976. Más de 600.000

registros que se pueden recuperar en formato ISBD, IBERMARC, formato personalizado del usuario. Lo distribuye Chadwyck Healey.

British national bibliography. Más de un millón y medio de registros de la bibliografía nacional británica visualizables y exportables en ISBD y UKMARC desde 1950 hasta ahora. CD-ROM realizado y distribuido por Chadwyck Healey

Bibliografia nazionale italiana. Más de 450.000 títulos en formato ISBD/Unimarc, editados desde 1958 hasta hoy.

Deutsche nationalbibliographie. Más de 500.000 títulos de la bibliografía nacional alemana, desde 1986 hasta la actualidad.

Catalogue de Belgique sur CD-ROM. Editado por la Universidad de Gent y la de Leuven, contiene 2.600.000 monografías disponibles en las 30 bibliotecas especializadas más importantes de las universidades e institutos científicos de Bélgica. También se ha añadido el catálogo colectivo de publicaciones periódicas.

Bibliographie nationale française. En dos CD-ROM se reúnen 800.000 registros desde 1970 hasta el presente Editado por Chadwyck Healey.

Bibliografia nacional portuguesa. 100.000 registros de documentos editados en Portugal desde 1980. Lo distribuye Chadwyck Healey.

## **1.6. Bases de datos bibliográficas y de libros en venta**

Alice-CD. Base de datos bibliográfica en la que se recogen 320.000 títulos publicados en lengua italiana, con más de 90.000 autores y 3.000 nombres y direcciones de editores, con 60.000 títulos out of print. Reúne los títulos contenidos en Italian Books in Print. Editado por Bowker.

Bibliodisc. Más de 400.000 libros publicados o distribuidos en Canadá.

Bibliografía general española. Contiene los registros bibliográficos de un millón de títulos publicados en castellano en España, Iberoamérica, y todo el mundo con fondos de 148 bibliotecas norteamericanas y europeas.

Bookbank. Información bibliográfica de libros editados en Europa con más de 600.000 libros y 22.000 editores. Incluye también títulos editados en Estados Unidos, disponibles en el Reino Unido. Editada por Whitaker. Se complementa con los CD-ROM: Bookbank with Thorpe y Bookbank OP.

Bookbank with Thorpe Plus. Toda la información de Bookbank, más los libros en lengua inglesa publicados en Australia, Nueva Zelanda e Islas del Pacífico. Editado por Whitaker.

Bookbank-OP. Incluye los libros publicados después de 1970 que dejaron de editarse a partir de 1991. Se actualiza anualmente con los títulos del Bookbank que han dejado de editarse en el año. Editado, al igual que los anteriores, por Whitaker.

Books in print plus. Más de millón y medio de libros americanos en venta, publicados por 52.000 editores. Cuenta también con: Books in print whit book reviews plus, que incluye más 250.000 reseñas de libros, extraídas de revistas especializadas y Books out print whit book review plus, con unos 700.000 títulos que han dejado de editarse desde 1979, aportando información sobre títulos alternativos, distribuidores de libros que ya no se editan, etc.

Bowker-Whitaker global books in print plus. Recoge dos millones de libros en lengua inglesa desde 1970, incluidos en otras bases de datos como Books in Print, Bookbank, International books in Print. Al igual que los tres anteriores está editado por Bowker Saur.

CD-LID. Contiene la base de datos de Libros españoles en venta. Incluye la base de datos Catalogo de Obras españolas en venta, creada por LID en colaboración con los editores, incorpora los títulos del ISBN, Ingram, Cove, Libri, Hethcote y Central Boekhuis.

Cumulative books index. Publicado por Wilson, recoge información actualizada de libros escritos en lengua inglesa publicados en todo el mundo desde 1982.

Children books in print. Libros infantiles editados en Estados Unidos. Distribuido por Bowker.

Electre-Biblio. Esta base corresponde a French Books in Print y contiene Les livres disponibles en France, con más de 380.000 registros de libros publicados por editores franceses en cualquier idioma o por editores extranjeros en francés. Proporciona siete formatos diferentes, entre ellos UNIMARC, LC-MARC y UK-MARC, lo que facilita su uso como sistema de catalogación por bibliotecarios. Esta realizado por Electre.

German books in print. Producida por Bowker, recoge 500.000 títulos de obras publicadas en Alemania, correspondientes por lo general al ISBN alemán. Los libros alemanes descatalogados han sido recogidos en German books out print, CD-ROM producido por Bowker, que recoge 250.000 títulos de libros impresos desde 1900 y que están descatalogados desde 1986.

Global books in print. Producida por Bowker y Whitaker, es una base de datos conjunta de estas dos importantes compañías con alrededor de 2.500.000 registros procedentes del Books in print y más de 750.000 de Bookbank. Estas dos grandes empresas ya colaboraron, hace más de 30 años, en la creación del sistema ISBN.

International guide to microform masters. Proporciona las colecciones en microforma de cerca de 200 bibliotecas a instituciones de investigación de USA; Canadá y Europa. Contiene un millón de entradas y datos técnicos sobre el tipo de película empleado, así como bibliográficos.

International books in print plus. Editada por Bowker Saur, incluye más de 250.000 títulos en lengua inglesa publicados fuera de Estados Unidos y el Reino Unido, con información sobre más de 6.000 editores y más de 1.000 distribuidores.

ISBN. Recoge más de 750.000 libros españoles en venta desde 1973 recopilados por la Agencia española del ISBN. Está realizado por Micronet. Existe la posibilidad de consultar esta base de datos vía Internet en la URL <<http://www.mcu.es/%20pic/spain/isbn.html>>

ISBN mexicano. Editado por Multiconsult, recoge los registros bibliográficos de libros publicados en México desde 1985.

Latbook. Esta base recoge libros latinoamericanos, editada por Fernando García Cambeiro, cuenta con la colaboración de diferentes bibliotecarios de países del ámbito hispanoamericano.

Libros en Venta en Hispanoamérica y España Plus. Cerca de 150.000 registros de obras en castellano editadas por alrededor de 5.000 editores de 36 países, sobre todo latinoamericanos, aunque también de otros como Estados Unidos o China. Editado por Boker-Saur

Russian books in print.- Editado por Bowker, contiene 200.000 referencias de libros publicados en Rusia y las antiguas repúblicas soviéticas desde 1989 hasta la actualidad.

Thorpe ROM. Editado por Whitaker, contiene información bibliográfica de libros escritos en lengua inglesa publicados en Australia, Nueva Zelanda y las Islas del Pacífico.

VLB Aktuell. Editada por Bowker Saur recoge desde 1986 más de 600.000 títulos editados en lengua alemana por unos 11. 000 editores. También incluye unos 240.000 registros de obras que los que dejaron de editarse.

Publishers international ISBN directory. Editado por Saur, ofrece datos de 350.000 editores de 210 países, con sus datos postales, número de ISBN, etc.

## **1.7. Catálogos de bibliotecas**

BiblioFile cataloging. Contiene el catálogo completo en formato MARC de la Biblioteca del Congreso. Permite exportar los registros a disco duro o disquete, facilitando la creación de bases de datos locales.

Bibliografía general española, siglo XV-1995. Un millón de títulos publicados en español en España, Ibero América y el resto del mundo, reunidos en los fondos de 148 bibliotecas norteamericanas y europeas. Distribuido por Chadwyck-Healey.

Bibliography of the hebrew books. Catálogo de la Jewish University Library de Jerusalén, con más de 9.000 títulos y 12.000 autores publicados en hebreo antes de 1960. Recoge libros impresos en caracteres hebreos desde 1473.

Bibliotecas sin fronteras. Contiene cerca de 250.000 registros del fondo americanista de las seis bibliotecas más importantes de España. Además incluye 15.000 artículos de revistas españolas y 3.500 tesis doctorales sobre América latina, leídas en diversas universidades europeas. Editada por Chadwyck-Healey.

Biblioteca infantil. Base de datos de literatura infantil del Centro Internacional del Libro Infantil y Juvenil de la Fundación Germán Sánchez Ruipérez, que es quien lo edita.

Bibliored. Catálogo colectivo de la red de bibliotecas públicas de México.

Biblioteca Nacional de Venezuela. Más de 200.000 registros procedentes del Instituto Autónomo Biblioteca Nacional y de otras bibliotecas de Venezuela.

Boston Spa books. 850.000 monografías recibidas en el British Library Document Supply Center. Desde 1980 lo distribuye la BLDSC.

British library general catalogue of printed books to 1975. Contiene los fondos bibliográficos de la Biblioteca Británica, con más de seis millones de registros con la mayor colección de libros anteriores a 1914 de todo el mundo. El 53% de la información es inglesa.

Catálogo colectivo bibliográfico. Más de 100.000 registros de las veinte bibliotecas universitarias de México.

CIRBIC-CSIC. Editado por Micronet, recoge el catálogo colectivo de las 80 bibliotecas del CSIC. Existe un fichero CIRBIC-libros, que recoge más de 600.000 monografías y otro, CIRBIC-revistas, que aglutina cerca de 37.000 títulos de publicaciones periódicas en los fondos del CSIC.

ECUADOR. Contiene más de 30.000 registros de la Biblioteca Central Pontificia de la Universidad Católica del Ecuador. Edita Multiconsult.

Edinburgh University Library. 300.000 registros que representan los fondos de la Universidad de Edimburgo.

LC MARC english. Editada por la Library Corporation recoge información de los títulos en lengua inglesa que posee la Biblioteca del Congreso, los registros

corresponden a todo tipo de material; libros, mapas, software, registros sonoros, material audiovisual.....

LC MARC foreign. Editado pro la Library Corporation agrupa los libros publicados en lenguas distintas al inglés, que posee la Biblioteca del Congreso.

LIBRUNAM. Base bibliográfica que contiene cerca de 500.000 registros catalográficos de los libros catalogados por la Biblioteca de la Universidad Autónoma de México.

Novum Regestrum. La base de datos más completa sobre fondo antiguo existente en las bibliotecas nacionales del ámbito latinoamericano, con información de 23 bibliotecas incluidas España y Portugal hasta 1850, aúna 180.000 registros. Editada por Chadwyck Healey.

The older books and most used nonbook cataloging collection. Contiene los registros más utilizados de libros antiguos y materiales no librarios editados con anterioridad a los últimos cinco años. Aproximadamente 1.500.000 registros en formato USMARC, del catálogo en línea de OCLC, que es quien lo edita.

The Recent books cataloging collection. Contiene más de 1.500.000 de registros en formato USMARC, de la Library of Congress, registrados por bibliotecas miembros de la OCLC, que es quien edita el disco, en los últimos cinco años.

REBIUN. Editado por Doc6 y Rebiun, mas de 200.000 registros de monografías y 28.000 de publicaciones periódicas, mas 17.000 de la colección Anticuaria, de la Universidad de Barcelona, en este catálogo colectivo de varias bibliotecas universitarias españolas que forman parte de la red Rebiun. Incluye la localización de los registros.

SERIUNAM. Editada por la Universidad Nacional Autónoma de México, contiene 30.000 registros de revistas existentes en las 163 bibliotecas de esta universidad.

VENEZUELA. Editada por Multiconsult, contiene 190.000 registros con la información de los libros en español catalogados por la Biblioteca Nacional de Venezuela.

WORLDCAT. Catalogo de OCLC, con 34.000.000 de registros, que describe documentos de todas las materias y materiales. Se da localización de cada registro. Lo edita OCLC.

Smithsonian ondisc. Casi medio millón de registros de bibliotecas de la Smithsonian Institution, de carácter multidisciplinar. Editado por G.K.Hall.

The Bodleian Library PRE -1920 catalogue of printed books. Oxford University Press, distribuye este casi millón y medio de registros multidisciplinares de la Bodleian Library, biblioteca fundada en 1598, ingresadas hasta 1920.

## **2. Documentos Web y otras informaciones accesibles vía Internet**

No cabe duda de que el fenómeno Internet ha provocado nuevos hábitos socio-culturales. La consulta de documentos en formato Web, el acceso a catálogos remotos o el uso de bases de datos a través de Internet son hábitos cada vez más arraigados entre los investigadores. Pero el gran problema de Internet es precisamente la multitud de documentos que están disponibles en la red, siendo necesario dedicar mucho tiempo a la búsqueda y selección de los Web de mayor interés. En el apartado que sigue a continuación se relacionan los lugares de mayor interés para quienes estudien cualquier ámbito de la Biblioteconomía.

---

### **2.1. Directories of resources on Biblioteconomía**

Biblioteconomía y Documentación en Internet <  
<http://www.ucm.es/BUCM/byd/05.htm>> La biblioteca de la Escuela de Biblioteconomía y Documentación de la Universidad Complutense ha realizado esta recopilación de recursos Web relacionados con la información y las bibliotecas. Incluye publicaciones periódicas electrónicas, bases de datos, literatura gris, enlaces a asociaciones y centros de formación, listas de discusión, etc. De interés porque reúne los recursos españoles más importantes.

BUBL: Library and Information Sciences <<http://link.bubl.ac.uk/lis>> Apartado dedicado a recopilar recursos sobre Biblioteconomía y Documentación, de entre los que realiza la escocesa universidad de Strathclyde. Utiliza la clasificación de Dewey para organizar los recursos, aunque también se puede acceder alfabéticamente o mediante un buscador. Ofrece un pequeño resumen de cada documento o enlace, así como el número de clasificación DDC correspondiente.

Internet per a bibliotecaris-documentalistes  
<<http://www.ub.es/div5/biblio/recursos.htm>> Recursos especializados en biblioteconomía, documentación y archivística, seleccionados por la E. U. Jordi Rubió i Balaguer de Barcelona. Incluye enlaces sobre la navegación en Internet y una serie de links especializados, divididos en directorios de recursos, escuelas de Biblioteconomía, bibliotecas y centros de información, distribuidores y productores de bases de datos, catálogos de bibliotecas y otras bases de datos bibliográficas, centros de suministro y préstamos interbibliotecario, archivos, editoriales, asociaciones, etc.

Library and Information Science Resources (IFLA) <<http://www.nlc-bnc.ca/ifla/ll/libdoc.htm>> Amplia selección de recursos accesibles vía Internet realizada por la IFLA, Federación Internacional de Asociaciones de Bibliotecarios y

Bibliotecas. Forma parte del conjunto de páginas web que ha realizado esta federación. Aquí se recopila una bibliografía de documentos electrónicos y se facilitan enlaces a directorios de recursos especializados en bibliotecas, así como a publicaciones electrónicas y a otros lugares relacionados.

Library and Information Science Resources (Library of Congress) <http://lcweb.loc.gov/global/library/> Agrupación de recursos sobre bibliotecas seleccionados por la Biblioteca del Congreso de Estados Unidos. Se incluyen categorías como recursos generales, bibliotecas nacionales, catálogos en línea, información y referencia, asociaciones profesionales, etc.

PICK: quality Internet resources in library and information science <http://www.aber.ac.uk/~tplwww/e/> Copiosa selección de recursos elaborada por la biblioteca Thomas Parry Library de Aberystwyth (Reino Unido). Es interesante la colección de documentos electrónicos a texto completo, así como sus recopilaciones de publicaciones electrónicas, recursos técnicos o enlaces con bibliotecas.

Thot <http://piramide.unizar.es/thot/thot.html> Selección de recursos sobre Biblioteconomía y Documentación recopiladas en el Área de Biblioteconomía y Documentación de la Universidad de Zaragoza. Divide los páginas de la selección en universidades, investigación, organizaciones, productos y servicios comerciales, publicaciones electrónicas, recopilaciones, bibliotecas y seleccionados. Facilita más de 150 enlaces.

Selección de recursos Internet sobre Biblioteconomía y Documentación <http://www.unileon.es/dp/dph/recursos/bibdoc/index.htm> Agrupación selectiva de enlaces sobre archivos, bibliotecas y documentación realizada por el Área de Biblioteconomía y Documentación de la Universidad de León. Divide sus páginas entre: guías y directorios, bibliotecas en Internet, bases de datos, publicaciones electrónicas, formación y profesión (universidades, directorios, asociaciones), editoriales y librerías, recursos técnicos (catalogación, clasificación, etc.) y acceso al documento.

## 2.2. Bibliotecas y catálogos

Bibliotecas españolas y de todo el mundo. <http://www.unileon.es/dp/abd/bibliot.htm> Directorio de bibliotecas con información en Internet. Principalmente reúne bibliotecas españolas, indicándose si el OPAC se puede consultar en línea o si, por el contrario, sólo aporta información de las bibliotecas. También facilita una completa relación de bibliotecas nacionales y de enlaces a directorios de bibliotecas de otros países. Presenta una relación muy completa, que suele estar actualizada.

Bibliotecas y centros de documentación en RedIRIS <http://www.rediris.es/recursos/bibliotecas/> Relación de bibliotecas y centros de información a lo que se puede acceder a través de la red de investigación

RedIRIS. Ofrece el enlace a las páginas web de la biblioteca, así como al catálogo, ya sea consultable a través de telnet o de WWW. Recientemente se han introducido en esta página nuevas informaciones sobre los fondos de la biblioteca, ubicación, etc.

EUROPAGATE <<http://olivo.csic.es/>> Pasarela para la consulta de Catálogos y Bases de Datos que soporten protocolo z39.50 o ISO/SR. Es fruto del proyecto del mismo nombre del programa Telematics for Libraries de la Unión Europea. Facilita las consultas simultáneas o independientes a una veintena de catálogos bibliotecarios europeos.

Gabriel Gateway to Europe's National Libraries <[Hytelnet](http://www.ucm.es/INET/hytelnet/html/start.html)> Mirror del famoso directorio de bibliotecas creado por Peter Scott. Al encontrarse la información en el servidor de la Universidad Complutense de Madrid la consulta se hace en menos tiempo. Esta versión incluye un sistema de búsqueda, aunque la información final siguen siendo las instrucciones para la consulta a través de telnet de los OPAC de bibliotecas.

WWW Accessible National Libraries of the World <<http://www.nlc-bnc.ca/ifla/II/natlibs.htm>> Directorio de bibliotecas nacionales con información web. Se encuentra dentro de las páginas elaboradas por la IFLA, donde se pueden encontrar otra serie de directorios similares. Solamente se da información postal y el enlace a las bibliotecas.

The WWW Library Directory < <http://www.llv.com/~msauers/libs/libs.html>> Directorio que incluye más de 1500 bibliotecas de todo el mundo accesibles vía world wide web. La búsqueda puede hacerse directamente o a partir del listado alfabético de países. Facilita además otros enlaces a otros recursos relacionados como empresas, instituciones o editores.

Web Cats: Library Catalogues on the World Wide Web <<http://library.usask.ca/hywebcat/>> Probablemente sea el directorio internacional más completo. Divide las bibliotecas por países, tipos de software de recuperación y tipología bibliotecaria. Completa al tradicional Hytelnet, directorio realizado por Peter Scott, el mismo autor de Web Cats, para facilitar el acceso telnet a distintas bibliotecas. El mayor problema de ambos directorios es que se centra demasiado en bibliotecas de Estados Unidos y Canadá.

### **2.3. Bases de datos**

Bases de datos de Fundesco <<http://www.fundesco.es/bases/bases.html>> Página de presentación de las distintas bases de datos que realiza Fundesco, entre las que destaca el catálogo de servicios ASCII, que contiene los servicios de información electrónica españoles accesibles en línea, en CD-ROM, en disquete y cinta magnética. Otras bases de datos son el catálogo de los fondos bibliográficos, la base de BBS (Bulletin Board Systems), el directorio de productos y servicios multimedia y el directorio de empresas e instituciones multimedia.

COMPLUDOC. <<http://www.ucm.es/bucm/inf/infodoc/>> Base de datos de publicaciones periódicas de Ciencias Sociales, entre las que se incluyen algunas revistas sobre Biblioteconomía y Documentación. Está realizada por la Facultad de Ciencias de la Información de la Universidad Complutense con la colaboración de alumnos de la Escuela de Biblioteconomía y Documentación de esta Universidad

Datatheke < <http://exlibris.usal.es/SFgate/dttlibre.htm>> Base de datos elaborada por la biblioteca de la Facultad de Traducción y Documentación de la Universidad de Salamanca. Para su realización se han vaciado los artículos sobre archivos, bibliotecas, documentación o temas afines publicados en las más de 250 revistas recibidas en la biblioteca.

Infobiba: Información y Bibliotecología Latinoamericana <<http://cuib.laborales.unam.mx/infobi.html>> Base de datos de pago elaborada por el Centro Universitario de Investigaciones Bibliotecológicas de la Universidad Nacional Autónoma de México. Se inicia en 1986 y en estos momentos contiene unos 6200 registros de publicaciones sobre humanidades, archivos, bibliotecas y temas relacionados.

ISBN < <http://www.mcu.es/pic/spain/ISBN.html>> Bases de datos bibliográfica producida por la Agencia Española del ISBN. Facilita información de más de 650.000 documentos, incluyendo todos los libros editados en España desde el año 1.966. Se actualiza periódicamente.

Library Literature <<http://www.hwwilson.com/liblit.html>> Base de datos especializada en Biblioteconomía y Documentación disponible hasta hace poco sólo en CD-ROM, pero ahora también accesible vía Internet, aunque mediante pago. Analiza más de 225 publicaciones periódicas y más de 600 libros y publicaciones de congresos.

Library Reference Center < <http://www.epnet.com/lrc.html>> Ebsco realiza esta base de datos de acceso gratuito que indiza y resume cerca de 30 revistas profesionales de Biblioteconomía y Documentación, sobre todo anglosajonas.

PCI: Periodical Contents Index <<http://pci.chadwyck.com/>> Repertorio internacional que analiza más de 1800 publicaciones de todo tipo de disciplinas, entre las que se encuentra Biblioteconomía y Documentación, representada por unas 50 revistas de todo el mundo. Esta base de datos contiene unos 7 millones de documentos y en las condiciones de suscripción presenta ciertas ventajas para las bibliotecas.

ReID: Base de datos de resúmenes de información y documentación. <<http://nóstromo.servitel.es/sisdoc/reid/inicio.HTM>> Base de datos que resume e indiza artículos sobre información y automatización de una veintena de revistas españolas, así como de actas de congresos celebrados en nuestro país. El análisis documental efectuado es exhaustivo, pudiendo recuperarse la información por varios campos. Desde el 1 de junio de 1997 el acceso a esta base es sólo

para usuarios registrados, lo cual se puede hacer por una cantidad no muy grande. No obstante, se puede consultar una versión demo totalmente operativa.

TESEO <<http://www.mec.es/teseo/>> Esta base de datos facilita información de las tesis doctorales leídas en las universidades españolas desde 1976. El acceso a la información puede realizarse de forma simple o utilizando operadores y modificadores de búsqueda. El resultado de la consulta mostrará una ficha con los datos principales de la tesis, más el resumen, descriptores y clasificación de la UNESCO.

UnCover <<http://uncweb.carl.org/>> UnCover incluye registros de artículos de unos 17.000 revistas multidisciplinarias, de entre las cuales unas 250 corresponden a temas relativos a la información. Su fondo alcanza 7 millones de artículos, publicados desde 1988. Es una base de datos de pago.

## 2.4. Listas de distribución

BIB-MED <<http://www.rediris.es/list/info/bib-med.html>> Lista alojada en RedIRIS destinada a la cooperación e intercambio de información de profesionales de bibliotecas médicas. Para obtener información habrá de enviarse un mensaje a BIB-MED-request@LISTSERV.REDIRIS.ES y para darse de alta el mensaje debe enviarse a lis-tserv@rediris.es indicando como texto del mensaje lo siguiente: subscribe bib-med Nombre Apellido, el subject (tema) del mensaje puede dejarse en blanco.

Biblio. Lista de discusión de bibliotecarios y documentalistas chilenos La dirección para suscribirse es: listserv@cobre.reuna.cl, dejando vacío el subject e indicando en el contenido del mensaje: subscribe biblio Nombre y apellidos.

Bibliomex-L. Lista mexicana orientada a la información y la cooperación bibliotecaria y temas relacionados. Para darse de alta hay que enviar un mensaje a: bibliomex-l-request@ccr.dsi.uanl.mx, sin poner nada en el "subject", e indicando en el cuerpo del mensaje: subscribe bibliomex-l Es posible enviar mensajes sin estar suscrito a la lista. En este caso la dirección es: bibliomex-l@ccr.dsi.uanl.mx

Educbiblio-l. Lista orientada a tratar temas sobre la educación en Biblioteconomía y Documentación. Se crea en México, aunque la temática es internacional. La dirección para suscribirse es: educbiblio-l-request@ccr.dsi.uanl.mx. Hay que enviar un mensaje a la dirección anterior, dejando el tema (subject) en blanco y poniendo como texto del mensaje: subscribe educbiblio-l. La lista permite enviar mensajes a personas no suscritas. Ello se hará a la dirección: educbiblio-l@ccr.dsi.uanl.mx.

FORO-L. Lista de discusión administrada desde la Universidad de Arizona y que trata de asuntos relativos a bibliotecas, sobre todo de México, Estados Unidos y Canadá. Para suscribir a se debe enviar un mensaje a: LIS-

TSERV@ARIZVM1.CCIT.ARIZONA.EDU, dejando el tema en blanco y poniendo en el cuerpo del mensaje: SUBSCRIBE FORO-L Nombre Apellidos.

Iwetel <<http://www.rediris.es/list/info/iwetel.html>> Foro de debate de profesionales de la bibliotecas y la documentación en general. Fue puesta en marcha en 1993 por los redactores de la revista Information World en Español. Es utilizada para solicitar información, petición de documentos, cooperación bibliotecaria, etc. Para suscribirse hay que enviar un mensaje a [listserv@listserv.rediris.es](mailto:listserv@listserv.rediris.es), indicando en el cuerpo del mensaje: subscribe iwetel Nombre Apellidos. Es una lista cerrada que sólo admite mensajes de los miembros, aunque los mensajes pueden ser consultados en <<http://chico.rediris.es/archives/iwetel.html>>

Listas de distribución (RedIRIS) <<http://ccr.dsi.uanl.mx/~ssouto/listas.html>> Página de RedIRIS destinada a informar sobre las listas de distribución alojadas en esta red científica. Se facilitan todos los datos para conocer qué listas españolas e internacionales existen y cómo suscribirse a las mismas. Posibilita la suscripción directa a muchas de las listas incluidas.

Listas hispanoamericanas de Bibliotecología  
<<http://ccr.dsi.uanl.mx/~SSOUTO/listas.html>> Descripción y condiciones de suscripción de las listas de discusión sobre Biblioteconomía que existen en América Latina. La información está realizada por Saúl Souto y permite enviar mensajes a casi todas las listas reunidas en esta página.

Liszt <<http://www.liszt.com/select/Humanities/Libraries/>>  
<[http://www.liszt.com/select/Humanities/Library\\_Science/](http://www.liszt.com/select/Humanities/Library_Science/)> Directorio multidisciplinar de listas de distribución. Presenta las principales características de cada una de ellas, indicando su temática y las instrucciones necesarias para suscribirse a las mismas. La primera de las URL aquí presentadas corresponde a las listas sobre bibliotecas recogidas en este directorio, mientras que la segunda reúne las específicas de Biblioteconomía y Documentación.

Mailbase <<http://www.mailbase.ac.uk/category/P.html>>  
<<http://www.mailbase.ac.uk/category/ZC.html>> Mailbase es un directorio de listas de discusión, entre otros servicios de Internet. Las direcciones que se presentan aquí reúnen las listas relativas a Biblioteconomía y bibliotecas, respectivamente. Ofrece una extensa información de cada lista, además de permitir la búsqueda en los archivos de las mismas.

Tile.net- Library <<http://tile.net/lists/library2.html>> Tile.net es un completo directorio de listas de distribución, news, servidores ftp, etc. En el apartado de bibliotecas recoge unas 375 listas, dándose de cada una datos como el número de participantes, el país de residencia y las direcciones para informarse y suscribirse.

## 2.5. Publicaciones electrónicas

### Libray-Oriented List and Electronic Serials.

<http://info.lib.uh.edu/liblists/liblists.htm> Directorio de listas de distribución y publicaciones electrónicas de interés para bibliotecas recopilado por la biblioteca de la Universidad de Houston. Es uno de lo más completo que existen, por lo menos el más exhaustivo de los especializados en Biblioteconomía.

NewJour. <http://gort.ucsd.edu/newjour/> Listado de más de 400 publicaciones electrónicas y boletines de noticias de todas las disciplinas. Es una relación completa, pero se echa en falta una clasificación temática.

Revistas electrónicas sobre Biblioteconomía y Documentación.  
[http://www.ugr.es/~felix/g7/g7\\_home.htm](http://www.ugr.es/~felix/g7/g7_home.htm) Repertorio realizado en 1996 por la Facultad de Biblioteconomía y Documentación de la Universidad de Granada. Bajo el concepto de revista electrónica incluye también muchas listas de discusión.

Se citan a continuación las direcciones de las principales revistas profesionales que se encuentran en formato electrónico. Algunas únicamente ofrecen información sobre la revista o los sumarios de la publicación, aunque en otras se facilita el texto íntegro de las mismas

Bulletin des Bibliothèques de France <http://www.enssib.fr/Enssib/bbf.htm>  
Cataloging & Classification Quarterly <http://ccq.libraries.psu.edu/ccq.html>  
Correo Bibliotecario <http://www.bcl.jcyl.es/correo/>  
Cuadernos de Documentación Multimedia  
<http://www.ucm.es/info/multidoc/revista/>  
Documentaliste <http://www.adbs.fr/adbs/prodserv/document/html/1som.htm>  
IFLA Journal <http://www.nlc-bnc.ca/ifla/V/iflaj/index.htm>  
International Information & Library Review  
<http://www.hbuk.co.uk/ap/journals/ir.htm>  
Investigación Bibliotecológica <http://Investigación Bibliotecológica>  
[http://cuib.laborales.unam.mx/publicaciones/revista/ca\\_rev.html](http://cuib.laborales.unam.mx/publicaciones/revista/ca_rev.html)  
Journal of Education for Library and Information Science  
<http://www.libraries.wayne.edu/LISP/jelis.html>  
Library Journal Digital <http://www.ljdigital.com/>  
Library & Information Science Research <http://www.lib.siu.edu/swen/iclc/lisr.htm>  
Library Trends <http://edfu.lis.uiuc.edu/puboff/catalog/trends/toc.html>  
LIBRES: Library and Information Science Research Electronic Journal  
<http://aztec.lib.utk.edu/libres/>  
Métodos de información <http://www.uv.es/cde/mei/>  
Revista Española de Documentación Científica  
<http://www.cindoc.csic.es/revind.htm>  
Revista General de Información y Documentación  
<http://www.eubd.ucm.es/publicaciones/revista.html>  
The Library Quarterly <http://www.gslis.ucla.edu/LIS/lq/>

## 2.6. Aspectos profesionales: asociaciones, formación, instituciones, etc.

ADBS <<http://www.adbs.fr/>> Excelente web de la asociación de los profesionales de la información franceses. Además de datos sobre la asociación y sus actividades, incluye documentos de interés profesional, así como los resúmenes de la revista Documentaliste.

American Library Association (ALA) <<http://www.ala.org/>> Páginas web de la asociación de bibliotecarios más importante del mundo. Ofrece copiosa información acerca de sus actividades, publicaciones, servicios, etc. Web interesante por la autoridad de esta asociación.

Directorio de correo electrónico de profesionales de documentación y bibliotecas en España <<http://www.cindoc.csic.es/prod/direc1.htm>> Relación de correos electrónicos de profesionales de la información españoles. Está elaborado por el CINDOC, basándose en los directorios ¿Quién es quién en información y documentación?

<<http://olivo.csic.es:4500/ALEPH/SPA/MAR/QEQ/MAR.QEQ/START?>> y el directorio IWE de profesionales de la información <<http://escher.upc.es/info/diriwe.htm>>

International Federation for Information and Documentation (FID) <<http://fid.conicyt.cl:8000/>> Páginas web de la Federación Internacional de Documentación, que cuenta con miembros de cerca de 90 países y que trabaja por la realización de políticas globales en el ámbito de la información.

International Federation of library Associations and Institutions (IFLA) <<http://www.nlc-bnc.ca/ifla/home.html>> Web de la federación más importante en materia de bibliotecas más importante del mundo. Informa de sus actividades, publicaciones, etc., aunque lo más destacable es la selección de recursos que realizan. Existe una copia (mirror) de estas páginas en Europa en <<http://ifla.inist.fr/>>

The Library Association <<http://www.la-hq.org.uk/>> Web de la asociación británica de profesionales de la información. Ofrece información sobre sus publicaciones, actividades asociativas y editoriales, proyectos, etc.

Professional Organizations in the Information Sciences <<http://witloof.sjsu.edu/peo/organizations.html>> Exhaustiva recopilación de asociaciones de bibliotecarios, documentalistas y profesionales de la información de todo el mundo. Apenas recoge asociaciones españolas.

Universidades españolas. <<http://www.unileon.es/dp/abd/ADAB/univer.htm>> Recopilación de enlaces a las Facultades o Escuelas Universitarias españolas que imparten la Diplomatura en Biblioteconomía y Documentación y/o la Licenciatura en Documentación.

World List of Departments and Schools of Information Studies  
<http://www.shef.ac.uk/~is/lecturer/tom2a.html> Relación geográfica de universidades de todo el mundo donde se imparten estudios sobre información y documentación. Es muy completa en el ámbito angloamericano, pero presenta bastantes carencias en el resto de los países.

## **Bibliografía**

- Abadal Falgueras, Ernest. Els serveis d'informació electrònica: què son i per a què ser-viexen. Barcelona: Universitat de Barcelona, 1997.
- Merlo Vega, José Antonio; Sorli Rojo, Ángela. Biblioteconomía y Documentación en Internet. Madrid: CINDOC, 1997.
- Educación y biblioteca. Biblioteca del bibliotecario (dossier). Mayo 96, n. 68, p. 27-63.

# **Capítulo 4. Sistemas de Recuperación de Información. Lenguajes de interrogación y operadores. Estrategias de búsqueda.**

## **Introducción a los Sistemas de Recuperación de información**

**Departamento de Información y Documentación.**

**<http://www.um.es/gtiweb/fjmm/sarisite>**

### **Antecedentes**

Las décadas de 1950 y 1960 ven el auge que alcanzan los estudios sobre comunicación, psicología, lógica, cibernetica, telecomunicaciones, teoría de sistemas, etc. al tiempo que una imparable multiplicación de los recursos informativos se da en todos los campos de la actividad humana. Los especialistas se centran en el problema de la búsqueda y recuperación de información, documental o no, lo que les lleva al comienzo del ciclo, a los principios de representación y clasificación de los documentos según su contenido semántico. *Information Retrieval* era un síntoma de un cambio más profundo, que llegó en 1966-68, con la introducción de la *Information Science* (Ciencia de la Información), como ciencia integradora de la teoría, proceso y práctica documental, con otras ciencias complementarias, como la cibernetica, la informática, la teoría de la información y la comunicación, etc.

### **Concepto de recuperación de información**

Cuando un usuario se plantea la necesidad de obtener nueva información sobre un asunto o materia de su interés, está manifestando una carencia, una situación irregular de sus estructuras mentales y cognitivas. Es un estado mental de incertidumbre que mueve al individuo a desarrollar una serie de acciones para salir de ese estado.

La respuesta a este tipo de situaciones es un conjunto de actividades que desarrolla el individuo para salir del estado anómalo, o para solucionar su problema de espacio, actividades que están íntimamente relacionadas con la adquisición de nueva información, y con el proceso comunicativo pertinente.

**La recuperación de información es el conjunto de tareas mediante las cuales el usuario localiza y accede a los recursos de información que son pertinentes para la resolución del problema planteado.** En estas tareas desempeñan un papel fundamental los lenguajes documentales, las técnicas de resumen, la descripción del objeto documental, etc.

**En principio, la recuperación de información engloba las acciones encaminadas a identificar, seleccionar y acceder a los recursos de información útiles al usuario.**

**El proceso de recuperación de información.** El proceso de recuperación de información engloba numerosas tareas, de las que la consulta de recursos de información electrónica resulta ser una más de ellas. Un proceso de recuperación, al que podríamos considerar "genérico" seguiría las siguientes fases:

1. Definición de las necesidades informativas del usuario.
2. Selección y ordenación de las fuentes a utilizar.
3. Traslación de las necesidades del usuario al lenguaje documental propio de la fuente a utilizar en cada caso. Es posible, además, encontrar fuentes en las que no se utilice ningún tipo de vocabulario controlado, en cuyo caso resultará necesario afinar el trabajo terminológico.
4. Traducción de la expresión de lenguaje documental al lenguaje de interrogación propio de cada sistema.
5. Ejecución de las expresiones del lenguaje de interrogación obtenidas.
6. Consulta de las respuestas obtenidas, para analizar su pertinencia o no a la cuestión planteada.
7. Replanteamiento, si procede, de las expresiones utilizadas, si los resultados obtenidos no son pertinentes.
8. Selección y obtención de los documentos que respondan a las necesidades manifestadas por el usuario.
9. Transmisión del resultado, preparado adecuadamente, al usuario.

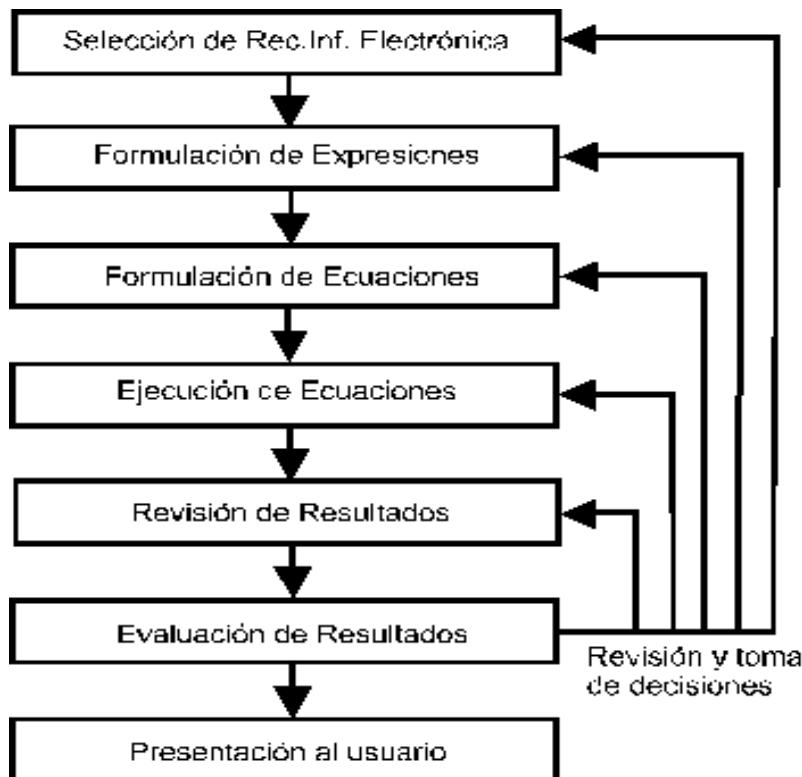


Fig. Proceso de recuperación en un entorno informático.

## ¿Para qué sirve un SRI?

Un **Sistema de Recuperación de la Información (SRI)** permite la recuperación de la información, previamente almacenada (claro está), por medio de la realización de una serie de consultas (“queries”) a los documentos contenidos en la base de datos. Esta serie de preguntas o interrogaciones se conceptúan como sentencias formales de expresión de necesidades de información, y suelen venir expresadas por medio de un lenguaje de interrogación.

Un documento es un objeto de datos, textual tradicionalmente, aunque la evolución tecnológica propicia la profusión de documentos multimedia, incorporándose fotografías, ilustraciones gráficas, vídeo animado, audio, etc.,

Un SRI debe soportar una serie de operaciones básicas sobre los documentos almacenados en el mismo, como son: introducción de nuevos documentos, modificación de los documentos almacenados y eliminación de los mismos.

Debemos también contar con algún método de localización de los documentos (o con varios generalmente), para presentárselos posteriormente al usuario.

## Clasificación de los SRI

La mayoría de los sistemas de información siguen dos modelos: booleano y de búsqueda de información por patrones de texto. Las interrogaciones a los sistemas de búsquedas por patrones de texto se llevan a cabo por medio de cadenas de caracteres. Los sistemas de patrones de textos son más utilizados comúnmente en pequeñas colecciones de datos y cuando hay que gestionar grandes volúmenes de documentos destacan mayoritariamente los sistemas booleanos.

En un sistema booleano, los documentos se encuentran representados por conjuntos de palabras clave, generalmente almacenadas en un fichero inverso. Un fichero inverso es una lista de palabras clave y de identificadores de los documentos en lo que éstas aparecen.

Las búsquedas booleanas consisten en expresiones de palabras claves conectadas con algún/nos operador/es lógico/os (AND, OR y NOT).

Al mismo tiempo que se han criticado estos sistemas, ha resultado complicado aumentar su efectividad en la recuperación de información. Sobre este modelo conceptual se han desarrollado algunas extensiones que se recogen bajo la denominación de modelo Booleano extendido.

Se pretende mejorar el rendimiento de los SRI usando la información procedente de la distribución estadística de los términos, ya que la frecuencia de aparición de

un término en un documento o conjunto de documentos se considera dato relevante al establecer una interrogación a la base de datos.

La distribución de frecuencias de un término se usa dentro del contexto de algunos modelos estadísticos, como es el caso del modelo de Espacio Vectorial, el modelo Probabilístico o el modelo conceptual Clustering.

Por medio del uso de estos modelos es posible asignar una probabilidad de importancia (un peso) a cada documento dentro de un conjunto de documentos recuperados para ser ordenados posteriormente según un cierto orden de importancia (pertinencia).

También resulta posible efectuar agrupaciones (clusters) de los documentos de la base de datos basándose en los términos que contienen y recuperar información desde estos grupos de documentos por medio de algoritmos de ranking.

### Estructuras de ficheros

Un fichero inverso es un fichero índice en cuya estructura de ítems encontraremos: palabra clave, Id-Doc, Id-Campo. Algunos sistemas incluyen información sobre la localización en el documento del párrafo y frase de los términos buscados.

Los ficheros de patrones de bits contienen hileras de dígitos binarios que representan a los documentos. Existen varias formas de construir estos patrones de bits.

Los grafos (o “redes”), son colecciones ordenadas de nodos conectados por arcos. Representan documentos de diversas formas y maneras. Un ejemplo es el grafo denominado red semántica, que representa las relaciones semánticas que se establecen en el texto, y que se pierden (a menudo), en otros sistemas de indización. Aunque son un campo interesante para el estudio, resultan difíciles de llevar a la práctica.

### **Operaciones de consulta**

Una de las operaciones de consulta más corriente es la operación denominada parsing, que consiste en la división de la interrogación (consulta) en sus elementos constituyentes. Las búsquedas booleanas deben ser divididas en sus correspondientes términos de indización o palabras clave y los operadores asociados a ellas para formular la expresión formal de la consulta. El conjunto de los identificadores de los documentos asociados con cada término de consulta es recuperado, y estos conjuntos son, entonces, combinados de acuerdo a los operadores booleanos.

La operación denominada reutilización (en inglés feedback), consiste en la reutilización de una búsqueda anteriormente efectuada. La información sobre el resultado de estas búsquedas es usada para formar parte de las consultas actuales; así, los términos de documentos relevantes encontrados en una consulta previa pueden añadirse a la consulta actual, y los términos correspondientes a documentos no relevantes pueden ser obviados. Hay creencia de que la reutilización de las búsquedas anteriores puede mejorar y potenciar las prestaciones de los SRI

### Operaciones sobre los términos

Con el concepto de stemming nos referimos a un proceso de “mezcla” (fusionando o combinando) de palabras relacionadas, reduciendo normalmente las palabras a su forma de raíz más común.

El truncamiento es una “mezcla” manual de términos usando caracteres especiales en la palabra, así que el término truncado formará múltiples palabras; en este caso nos referimos a las operaciones de localización de términos con una raíz común.

Otra asociación de términos relacionados la constituye un tesauro, el cual, nos va a ofrecer una lista de términos, sus términos sinónimos y las relaciones semánticas mantenidas entre los términos del mismo.

La lista de palabras vacías es una relación de términos considerados como valores no indizables, usados para eliminar potenciales términos de indización.

En cuanto a la ponderación de términos, a éstos se les puede asignar un valor numérico basado en su distribución estadística, o sea, la frecuencia con la que los términos aparecen en documentos, colecciones de documentos, o en subconjuntos de colecciones de documentos, tales como documentos considerados relevantes en una búsqueda (pregunta).

### Operaciones sobre los documentos

A los documentos añadidos a una BD se les asigna un identificador único; los documentos se estructuran en campos y los mismos deben ser introducidos dentro de identificadores de campos y conjuntos de términos. A veces queremos desenmascarar ciertos campos para buscarlos y mostrarlos. Otra operación es la ordenación de los documentos recuperados por algún campo determinado.

La operación de mostrar incluye tanto a la salida impresa de los documentos como a su visualización en la pantalla del ordenador. El diseño del interfaz de usuario de un SRI resulta de carácter vital, como en otro tipo de sistemas de información, para conseguir un uso efectivo del mismo.

De la información procedente de la distribución de frecuencias de los términos, es posible asignar una probabilidad de relevancia a cada documento dentro de un conjunto recuperado, permitiendo que los documentos recuperados sean organizados en orden a esta probable relevancia.

La información de la distribución de frecuencias de los términos puede ser usada para agrupar documentos similares en un espacio documental, por medio de las técnicas de clustering.

## **Evaluación de los SRI**

La eficacia en la ejecución es medida por el tiempo que se toma un sistema o una parte de un sistema para realizar una operación. Este parámetro ha sido siempre la preocupación principal en un SRI, especialmente desde que muchos de ellos son interactivos, y un largo tiempo de recuperación interfiere con la utilidad del sistema.

La eficiencia del almacenamiento es medida por el número de bytes que se precisan para almacenar los datos. El espacio general, una medida común de medir la eficacia del almacenamiento, es la razón del tamaño del índice de los ficheros además del tamaño de los archivos del documento sobre el tamaño de los archivos del documento. Las ratios del espacio general que oscilan entre los valores 1,5 y 3 son típicas de los SRI basados en los ficheros inversos.

De forma tradicional se ha conferido mucha importancia a la efectividad de la recuperación normalmente basada en la relevancia de los documentos, lo cual ha representado un problema ya que medir la relevancia es un proceso subjetivo y sin confianza. Muchas medidas de la efectividad de la recuperación han sido propuestas. Las más empleadas, de forma general, son las conocidas como *rellamada* y *precisión*.

### Exhaustividad y Precisión

Precisión es la ratio del número de documentos relevantes recuperados, sobre el número total de documentos recuperados. El rango de valores de ambas ratios, está comprendido entre 0 y 1.

Exhaustividad es la ratio de documentos relevantes recuperados en una búsqueda dada, sobre el número de documentos relevantes para esa búsqueda en la base de datos. Excepto para tests realizados sobre pequeñas colecciones, este denominador es generalmente desconocido y debe ser estimado por muestreo o por otros métodos.

### Medidas propuestas por Lancaster

Los criterios indicados anteriormente no resultan los únicos que se han propuesto, de hecho, casi desde el desarrollo de los primeros sistemas se vienen proponiendo otra serie de criterios, tal como es la propuesta de Lancaster quien establece seis criterios:

- Cobertura
- Exhaustividad
- Precisión
- Tiempo de Respuesta
- Esfuerzo del usuario
- Formato

Aunque estos criterios tienen casi tres décadas de antigüedad, algunos de ellos continúan perfectamente vigentes hoy en día e incluso hay que apuesta por su aplicación para evaluar las prestaciones de los distintos motores de búsqueda en Internet.

### Medidas propuestas para World Wide Web

Composición de los índices: la composición de los índices afecta de forma muy directa a la calidad de la recuperación de información.

Destacan tres componentes importantes: Cobertura, Frecuencia de Actualización y Porción de página Web indexada (título, título y primeros párrafos, página completa, etiquetas meta). Las magnitudes de cada motor dependerán del  $H_w$  y  $S_w$  dedicado. De otro lado, el que el índice sea muy extenso no implica calidad (el ejemplo lo tenemos en Yahoo) y, de otro lado, el poseer un valor muy alto en este parámetro tampoco implica altos niveles en los otros tres.

Capacidades de búsqueda: un motor ha de poseer operadores booleanos, búsquedas por expresiones literales, truncamiento de los términos y facilidades de acotar una búsqueda en un determinado campo. De hecho, este conjunto de prestaciones comienzan a considerarse básicas, destacando (negativamente) motores tan famosos como Lycos o Excite que no incluyen la búsqueda por frase literal.

Ejecución de la recuperación de información: suele medirse con base en tres parámetros: Precisión, Exhaustividad y Tiempo de Respuesta. Estas medidas, no obstante, incorporan muchas dosis de subjetividad a la hora de determinar cuándo un resultado es relevante o no.

Esfuerzo del usuario: la Documentación y el Interfase son elementos a considerar en este apartado y suelen tener un aceptable nivel. De hecho, este parámetro es muy importante porque un usuario no va a hacer uso de un motor si no se

encuentra cómo con su interfase, si no localiza fácilmente la documentación que indique cómo emplearlo y si no la comprende.

### **Lenguajes de interrogación y operadores**

Es obvio, que la flexibilidad y variedad del lenguaje natural crea serias dificultades para el manejo automatizado de la información. Para solucionar este problema se utilizan tesauros que permiten traducir información del lenguaje natural de los documentos y usuarios a un lenguaje controlado del sistema. Por ejemplo, en la Base de Datos Medline es posible utilizar el índice Mesh (Medical Subject Subheadings) el cual brinda un listado de términos controlados ó palabras clave utilizados en el campo de descriptores de la base de datos. Mesh es el vocabulario controlado creado por la Biblioteca Nacional de Medicina de Estados Unidos y la utilización de sus términos, permite realizar una búsqueda más precisa. El uso de descriptores aumenta la precisión pero confiar sólo en ellos puede reducir la recuperación, por lo tanto se aumenta la recuperación por usar tanto el descriptor como la frase de uso común ó término no controlado.

La otra parte de la adquisición del conocimiento está orientada a que las Bases de Datos proporcionen un modelo que le permita al usuario construir una plantilla de búsqueda conforme a las expectativas de dicho usuario. Los elementos considerados relevantes por un determinado usuario pueden variar, pero una plantilla que guíe la estrategia de búsqueda del profesional puede ser de utilidad si está relacionada con las preguntas que se hace el usuario a la hora de llenar esta plantilla.

En una sesión de trabajo del sistema, el primer paso es la generación de la plantilla básica de búsqueda. En este paso se define en primer término el tipo de documentos que le interesa recuperar; luego se define el ó los lenguajes de interés para la consulta; el tipo de fuente a la que desea consultar: motores de búsqueda (Google, Yahoo, Altavista), bases de datos (Medline, Lilacs, Cochrane), o revistas especializadas.

**Se puede definir a un lenguaje de interrogación como un conjunto de órdenes, operadores y estructuras que, organizados según unas normas lógicas, permiten la consulta de fuentes y recursos de información electrónica. El resultado de la combinación de estos elementos, siguiendo las normas establecidas, es una expresión, a la que se identifica con el nombre "ecuación", capaz de interrogar el contenido de la fuente de información.** Esta definición del lenguaje de interrogación y de sus componentes puede encontrarse en la norma ISO 8777-1988.

Las normas lógicas que rigen un lenguaje de interrogación responden a cuestiones relacionadas con la coordinación de los elementos, es decir, con la formulación de ecuaciones. Estas normas (a modo de sintaxis) especificarán el orden de los elementos, la disposición de las estructuras, sus posibilidades

combinatorias, las prioridades en la ejecución, y todo tipo de posibles funciones. Las órdenes serán aquellas palabras o abreviaturas que le indicarán al sistema las acciones a ejecutar (buscar la expresión, mostrar los documentos o registros resultantes, consultar el tesauro o los ficheros inversos, ejecutar un perfil de usuario, etc.).

Sin embargo, no todos los lenguajes de interrogación utilizan las mismas palabras como órdenes, aunque las órdenes ejecuten las mismas funciones. Existen intentos para homogeneizar la interrogación de las bases de datos, como el lenguaje CCL (Common Command Language) promovido por la Unión Europea, que aún no han alcanzado el objetivo para el que fueron desarrollados. A este panorama se une la proliferación de interfaces gráficos de usuario, que sustituyen a las órdenes y la sintaxis tradicional, dejando al usuario (si éste lo desea) sólo la labor de introducir los términos y los operadores que expresan las relaciones existentes entre ellos.

Uno de los puntos a considerar en el proceso de BRI, es el siguiente: la utilización de operadores lógicos aumenta la precisión en la búsqueda, pero la mayoría de los usuarios no los conoce o no sabe utilizarlos adecuadamente, y por ende no los aplica, dando como resultado búsquedas ineficientes.

En un lenguaje de interrogación, los operadores son los encargados de expresar las relaciones que mantienen entre sí los términos que definen (más adecuado sería decir que pueden definir) las necesidades informativas del usuario. Pueden distinguirse diferentes tipos de operadores, que se analizan a continuación.

### **Operadores lógicos (o booleanos)**

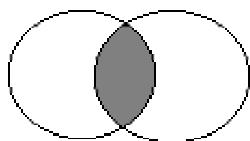
Los operadores lógicos, llamados booleanos en honor a George Boole, precursor de la lógica simbólica y del álgebra de conjuntos, son los más utilizados en numerosos sistemas. El principio que rige la utilización de este tipo de operadores es que las relaciones entre conceptos pueden expresarse como relaciones entre conjuntos. Las ecuaciones de búsqueda pueden transformarse en ecuaciones matemáticas, que ejecutan operaciones sobre los conjuntos, lo que da como resultado otro conjunto. Los tres operadores básicos son:

1. Operador suma/unión (generalmente identificado como O/OR)
2. Operador producto/intersección (identificado como Y/AND)
3. Operador resta/negación (identificado como NO/NOT).

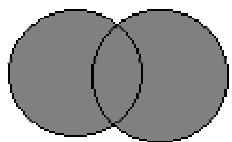
A su vez estos operadores pueden combinarse entre si, generando operaciones más complejas.

Independientemente de su potencia, no debemos obviar los problemas que presentan los operadores booleanos. En primer lugar, siempre se plantean en términos de absoluto (presente/ausente), sin considerar el peso específico del término en el contexto. En segundo lugar, exigen un alto valor de precisión en los

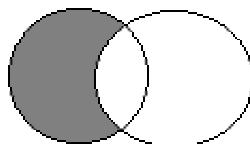
términos utilizados. Por último requieren claridad en la composición de las expresiones a buscar.



**AND/Y**  
Producto lógico



**OR/O**  
Suma Lógica



**NOT/NO**  
Resta lógica

### **Los tres operadores booleanos básicos.**

#### **Operadores posicionales.**

La utilización de operadores posicionales pretende superar algunas de las limitaciones que ofrecen los operadores booleanos. Toman como punto de partida la consideración del valor del término dentro del contexto, es decir, que la posición de ese término en relación con otros, o dentro del propio registro, es significativa para valorar su pertinencia a los objetivos buscados. Los operadores posicionales pueden dividirse en dos tipos:

➤ *Posicionales absolutos.*

Son aquellos que permiten buscar un término en un lugar dado del documento o registro. Por regla general, son operadores de campo, es decir, permiten al usuario fijar en qué campo o campos presentes en la estructura de base de datos debe aparecer el término buscado. La presencia del término en un campo dado (por ejemplo, en el campo título), puede ser una garantía de la adecuación del documento a los objetivos, en la mayor parte de las situaciones.

➤ *Posicionales relativos.*

También llamados de proximidad; se trata de operadores que permiten establecer la posición de un término respecto a otro dado. Se considera que la cercanía entre los dos términos puede reflejar una íntima relación entre los conceptos reflejados por los mismos. Estos operadores permiten definir el nivel de proximidad entre los términos (mismo campo, línea, frase, número de términos significativos que los separa, etc.).

## **Operadores de comparación.**

Especifican el rango de búsqueda, fijando unos límites para la misma. Estos límites pueden ser tanto numéricos como alfabéticos, correspondiendo los operadores a formas del tipo "mayor que", "menor o igual que". Se utilizan principalmente en documentos que pueden contener datos numéricos.

## **Operadores de truncamiento.**

Pueden darse situaciones en las cuales sea necesario utilizar no un término simple, sino también sus derivados, fijados por prefijación o sufijación, mínimas variantes léxicas, etc. Para facilitar la búsqueda de este tipo se han introducido operadores de truncamiento, a los que también se llama máscaras. Se trata de operadores (normalmente símbolos como \*, \$), cuya presencia puede sustituir a un carácter o a un conjunto de caracteres, situados a la izquierda, dentro o a la derecha del término en cuestión.

En los actuales sistemas de recuperación de información es posible encontrar todos estos tipos de operadores, que pueden combinarse entre sí, permitiendo crear ecuaciones complejas que reflejan con bastante precisión los conceptos y sus relaciones. La combinación de los operadores debe respetar un conjunto de reglas, básicas en todos los sistemas, que establecen las prioridades y formas de ejecución de ecuaciones complejas, cuando éstas combinan más de dos conceptos. En primer lugar, los sistemas tienden a resolver, a ejecutar en primer lugar, aquellas expresiones que se relacionan utilizando el operador más restrictivo o prioritario. Por ejemplo, un operador posicional absoluto posee un nivel de restricción (una prioridad) mayor que un operador booleano, lo que significa que el sistema ejecutará antes la expresión cuyo operador es el posicional absoluto, combinando posteriormente el resultado con el operador booleano y su término relacionado. Sin embargo, pueden darse expresiones en las cuales sea necesario variar estas prioridades, y ordenar al sistema que ejecute en primer lugar expresiones con operadores de menor nivel de restricción, relacionando luego su resultado con términos a través de operadores más restrictivos. Para estas situaciones, se utilizan paréntesis, los cuales engloban a las expresiones que deben ejecutarse en primer lugar, independientemente de las prioridades fijadas por el sistema. La utilización de expresiones entre paréntesis hace posible, por ejemplo, que el resultado de una expresión con un operador booleano pueda ser combinada con un operador posicional absoluto. Además, los paréntesis pueden anidarse, resolviéndose las ecuaciones planteadas desde dentro hacia fuera, de la misma forma que las igualdades y polinomios matemáticos.

En cuanto a los lenguajes para las búsquedas de información en la Web, resultan particularmente significativas las actuales iniciativas que pretenden automatizar el proceso de categorización y relevamiento de la información. Entre las más importantes se encuentran la Web Semántica y la Minería de Datos.

La Web Semántica es una propuesta de Tim Berners-Lee, inventor del WWW, cuyo objetivo es que los datos de la Web estén definidos mediante un lenguaje que pueda ser utilizado por las máquinas sin intervención humana. La Web Semántica pretende construir una estructura de metadatos, es decir datos sobre el significado de los datos almacenados, e incluirlos en los documentos de forma que sean identificables y comprensibles por las máquinas. La Web semántica codifica de forma automática los datos, mediante la especificación de un conjunto de conceptos u Ontologías. Las Ontologías se expresan mediante lenguajes de representación como el Resource Description Framework (RDF) que, en la Web semántica, se construyen sobre XML.

A diferencia de la Web Semántica la Minería de Datos pretende extraer los metadatos (patrones) a partir de información textual no necesariamente estructurada. Los sistemas de Minería de datos permiten, a partir del análisis léxico de los textos, la construcción automática de estructuras de clasificación y categorización. (Dürsteler, 2002)

Aún considerando que estas iniciativas son muy promisorias, continúan existiendo búsquedas en la web que no resultan fructíferas ni satisfactorias, en parte porque los buscadores actuales no han conseguido implementar exhaustivamente la tecnología disponible, o tal vez porque las expectativas de mejoría en las búsquedas deban volver la atención hacia la interacción del humano con el sistema.

### **Estrategias de búsqueda.**

El proceso de recuperación, extracción e integración de la información debe responder a una estrategia de búsqueda validada por el usuario, y es enviada a distintos tipos de fuentes y las respuestas de cada una de ellas, se presentan en una única respuesta al usuario. Las estrategias de búsqueda suponen inicialmente el desarrollo de un proceso de adquisición de conocimiento y su empleo en situaciones concretas, como, así también, la posibilidad de establecer mecanismos de actualización de dichas estrategias en la medida que los usuarios sean capaces de discriminar la información relevante de la que no es relevante. O lo que es lo mismo, que el sistema "aprenda" a mostrar la información que puede ser relevante para una búsqueda, basándose en un posible estudio estadístico de la relevancia de los campos en búsquedas calificadas como "éxito" por los usuarios. El proceso de aprendizaje de las estrategias de búsqueda tendrá un importante componente subjetivo, y en principio convendrá circunscribir el espacio de búsqueda al ámbito de publicaciones reconocidamente calificadas.

El término "estrategia", en lo que se refiere a la consulta de bases de datos, ha servido para identificar diferentes enfoques y conceptos, que engloban desde la visión general del proceso hasta la formulación de ecuaciones individuales.

*La estrategia debe ser un plan ideal de interrogación de la base de datos que incluya:*

- el objetivo de la búsqueda
- el plan general
- plan específico de operación.

El objetivo de la búsqueda se obtiene identificando qué tipo de información se necesita y sus características.

Una vez definido el objetivo, debe establecerse un plan general de operación, que incluya:

- una selección de la base o bases de datos a consultar
- las primeras aproximaciones a los términos a utilizar en las ecuaciones
- las posibles relaciones lógicas.

El plan específico de operación se pone en marcha una vez obtenidos los resultados del anterior, y debe:

- formular ecuaciones y utilizar términos con el mayor grado de precisión
- establecer una secuencia lógica con todo ello
- redefinirlo si es preciso.

Independientemente de ambos planes, resulta necesario conocer con anterioridad la respuesta a varias cuestiones que afectan a la interrogación de la base de datos, tales como el contenido y alcance de la base de datos, costo de consulta, lenguaje y operadores a utilizar durante las consultas, límites preestablecidos (por el usuario o el sistema); todas ellas afectan y modifican el enfoque del interrogador.

### Tipos de estrategias.

En el momento actual se considera que parece más adecuado utilizar el término estrategia para identificar el plan general de búsqueda. No existe una única ni perfecta aproximación a las estrategias de interrogación de bases de datos. En la mayor parte de las ocasiones esto depende de la experiencia del usuario y de la calidad del contenido de los registros existentes en la base de datos, especialmente en lo que corresponde a su control terminológico. La formulación de una estrategia depende, en gran medida, de la formación, intuición y experiencia del usuario. Tomando en consideración la intención del interrogador, la bibliografía señala que pueden existir varios tipos principales de búsqueda, que pueden clasificarse en dos grandes grupos, aunque puedan darse situaciones en las que estos tipos de búsqueda se combinen.

#### 1. Categorización por objetivo:

- Búsqueda de elemento conocido: se trata de búsquedas en las cuales el interrogador sabe cual será la respuesta, cuyo contenido, por ejemplo, utiliza para

completar una referencia bibliográfica, o utiliza un número de inventario para comprobar el contenido de un registro dado.

- Búsqueda de información específica: el interrogador busca una información específica dada, generalmente sobre un tema concreto y limitado, como trabajos publicados en un año o por un autor.

- Búsqueda de información general: intenta buscar la información sobre una materia o asunto, de forma general, que obtenga una visión global del estado de la misma.

- Exploración de la base de datos: se trata de conocer qué tipos de información y/o documentos se encuentran almacenados en la base de datos, a qué pueden responder y cómo pueden utilizarse.

## 2. *Categorización por plan de operación:*

- Búsqueda directa: se trata de una aproximación rápida, en la que se intenta resolver el problema con la formulación de una única consulta. Como puede deducirse, resulta difícil obtener buenos resultados con la misma.

- Búsqueda "breve": es una evolución de la anterior, en la que se trata de recuperar unos registros significativos entre un gran número obtenido tras una sola ecuación.

- Ampliación: comienza con ecuaciones muy restrictivas, que ofrezcan documentos pertinentes. Tras analizar la respuesta, el usuario puede ampliar o expandir las ecuaciones de búsqueda hasta recuperar toda la información existente. Puede ofrecer problemas si la ecuación inicial no es adecuada.

- Restricción: opuesta a la anterior, formula ecuaciones que ofrecen resultados muy amplios, para posteriormente utilizar ecuaciones más restrictivas, hasta delimitar los documentos pertinentes.

## **La exploración como mecanismo de recuperación**

Las limitaciones inherentes al proceso de recuperación mediante ecuaciones han conducido a experimentar otras maneras de realizar el proceso de búsqueda. Una de las más utilizadas es aquella que utiliza la exploración, es decir, el acceso a los documentos mediante técnicas de visualización de parte de su contenido que puede ser relevante, y la posterior asociación con otros documentos de perfil similar. Esta capacidad de exploración debería ser fundamental en los sistemas.

El usuario accede a un listado o enumeración de elementos descriptivos, y mediante un proceso de selección de elementos, va centrando el objetivo de su búsqueda. Los criterios utilizados por el usuario se basan en la deducción y la asociación de conceptos frente a la lógica de conjuntos que se plantea en un sistema de ecuaciones. Este tipo de representación es más adecuada para reflejar

las múltiples variaciones que un concepto puede tener para un usuario individual. En cambio, la utilización de la exploración suele realizarse en entornos en los cuales el usuario no posee una idea clara de cual debería ser la mejor táctica para aproximarse a la información que precisa.

Por lo tanto, la cuestión clave a considerar en un sistema de exploración es combinar las ideas y esquemas del usuario con el esquema de organización de la información que ofrece el sistema. Esta es la aproximación que pretenden desarrollar los enfoques cognitivos, poniendo su énfasis en el intermediario que debe existir entre el modelo del usuario y el modelo del sistema.

### **Estrategias de búsqueda en Internet**

Como nos hemos podido darnos cuenta, buscar en Internet es una tarea difícil por la gran cantidad de información, la ausencia de clasificaciones consistentes, las diferentes herramientas de búsqueda, los diversos modos de buscar, las presiones comerciales, etc.

Para conseguir llevar a cabo una buena tarea de búsqueda son necesarios tres elementos fundamentales:

- diseñar un perfil de búsqueda adecuado; es decir hay que tener la habilidad de emparejar con exactitud los términos que se utilizan en el perfil de búsqueda con los que se encuentran en los documentos cuya información se quiere encontrar
- conocer el tamaño y contenido de la base de datos que se elige, y
- conocer las posibilidades que ofrecen las diferentes herramientas de búsqueda para recuperar los contenidos de su base de datos.

La siguiente estrategia general puede resultar recomendable en muchos casos, pero para cada necesidad informativa deberemos prestar atención a sus características concretas y conducir la búsqueda en función de ellas:

- Antes de describir las diferentes herramientas de búsqueda que están disponibles es necesario analizar el tema antes de comenzar. Hay que intentar *formular correctamente* lo que queremos saber y traducir esta pregunta a un perfil de búsqueda. En Internet no hay *vocabulario controlado*, por lo que habrá que buscar palabras clave y cuantos sinónimos sean posibles para poder llevar a cabo la búsqueda correctamente. Además habrá que establecer relaciones entre las palabras, por si es posible llevar a cabo una búsqueda con operadores booleanos.
- **Si se pretende hacer una navegación para ver lo que hay sobre un tema sobre el que no se sabe muy bien lo que puede haber lo mejor es utilizar** alguno de los motores sencillos de utilizar como Google o FastSearch. Son herramientas útiles que permiten obtener rápidamente una información que suele dar una primera respuesta a nuestra necesidad informativa. Otra opción es recurrir a algún metamotor de búsqueda como

**Ixquick.** Esta última opción permite obtener una primera visión de lo que se está buscando de manera rápida y con las bases de datos de varios motores a la vez. Teniendo en cuenta que ninguna de las bases de datos es totalmente coincidente esta opción permite mayor exhaustividad.

- **Si se conoce bien lo que se quiere buscar lo mejor es utilizar los motores específicos y/o las opciones avanzadas de los principales motores de búsqueda.** Si hemos empezado por uno de los motores anteriores y hemos obtenido demasiados enlaces y no podemos explorarlos todos y conocemos bien el tema sobre el que estamos buscando información lo mejor es hacer una búsqueda lo más específica posible con buscadores específicos o con todas las potencialidades de la búsqueda avanzada de alguno de los grandes buscadores como **Altavista, Northern Light** u otros. En el caso de necesidades específicas siempre puede resultar necesario recurrir a **bases de datos específicas** que solo son accesibles utilizando las **herramientas** que permiten acceder a esta parte de Internet.
- **Utilizar los directorios de materias.** Si en los pasos anteriores se han obtenido demasiados enlaces como para buscar la información que responda a la pregunta de partida los **directorios por materias** permiten un acercamiento a recursos seleccionados desde lo más amplio a lo más específico por grandes áreas temáticas. Ahora bien sus bases de datos no tienen demasiados recursos y si no tienen motor para buscar en su propia base de datos desplazarse por sus clasificaciones temáticas no siempre es una tarea fácil.
- Si ninguno de los métodos anteriores nos conduce al éxito una última opción consiste en buscar la ayuda de un experto participando en lista de interés o buscando la información en grupos de noticias. Para saber si hay alguna lista que discuta el tema de nuestra incumbencia - en ellas con frecuencia participan verdaderos expertos en la materia - lo mejor es recurrir a directorios que nos pueden ayudar a localizarla como los que están accesibles a través de un recurso como **E-Mail Discussion Groups**. Para buscar los temas que se discuten en los grupos de noticias se puede recurrir a un motor de búsqueda como **Deja.com**.

### **Recomendaciones para búsquedas en Internet.**

De manera resumida, para las búsquedas de información en Internet, se puede sugerir una estrategia básica que incluye los siguientes pasos:

- Identifique los conceptos importantes de su búsqueda.
- Identifique los términos de búsqueda (keywords) para describir estos conceptos.
- Considere sinónimos, términos afines y variaciones de los términos (como plurales).
- Lea las instrucciones de ayuda en la página del buscador o directorio.
- Prepare la búsqueda lógica (aprenda la lógica simple de búsqueda booleana).

- Afine la búsqueda.
  - Hágala general (con OR o sinónimos)
  - Hágala más específica (con AND, operadores de proximidad, términos más precisos).
- Intente realizar la búsqueda en más de un buscador o directorio.

### **Revisión y análisis de resultados.**

Como es conocido, Internet es un medio de autopublicación, y este hecho debemos tenerlo siempre presente cuando se inicia un proceso de búsqueda en este medio. La calidad de la información que allí se encuentra, puede variar muchísimo, pues, generalmente, no se revisa por expertos antes de ser publicada. También hay que tener en cuenta que las páginas de Internet cambian con el tiempo de acuerdo con los intereses de los que las mantienen; así, nos podemos encontrar con documentos que ofrecen información elaborada por verdaderos especialistas, y con otros que se deben a la contribución de aficionados o de otras personas que publican por entretenimiento.

Por consiguiente, es de vital importancia, el proceso de evaluación de los resultados del proceso de búsqueda en Internet.

*El resultado de la ejecución de una ecuación de búsqueda es un conjunto de documentos que cumplen las condiciones expresadas en la ecuación.* Se trata, a su vez, de un subconjunto del conjunto total de documentos existentes en el recurso o fuente de información consultada. Sin embargo, puede darse el caso de que la respuesta sea un número excesivamente elevado de documentos, o un número mínimo. Por otra parte, los documentos resultantes responden a la lógica y a las condiciones expresadas en la ecuación de búsqueda, lo cual no supone, como ya se ha señalado, que sean pertinentes a las necesidades del usuario. En realidad, es posible ejecutar ecuaciones perfectas, desde un punto de vista funcional (operadores, términos, etc.), sin que los documentos resultantes reúnan las características que los harían deseables para el usuario.

*Para superar esta posible distorsión en los resultados es necesario valorar y evaluar la respuesta a las ecuaciones planteadas.* La primera modificación a realizar en la formulación de las ecuaciones afecta al número de respuestas obtenidas. En el caso de un excesivo número, se utilizan técnicas de restricción, mediante la introducción de términos más específicos, desechar términos generalistas, o limitar los truncamientos. En el caso de un número muy reducido, las acciones a tomar son las contrarias: utilizar términos más generales, incluyendo derivados y relacionados, limitar los operadores más restrictivos, añadir truncamientos, etc. Si se da la situación de ecuaciones correctas funcionalmente, pero sin respuesta adecuada, sería necesario replantear el proceso de recuperación, especialmente en la utilización de los lenguajes documentales y en la selección de fuentes.

Los resultados de una búsqueda se pueden valorar cuantitativamente utilizando dos parámetros, que son la **exhaustividad** y la **precisión**. Estos parámetros combinan el número de documentos pertinentes y no pertinentes, y recuperados y no recuperados, según las siguientes fórmulas:

	Pertinentes	No pertinentes
Recuperados	a	b
No recuperados	c	d

La tasa de **exhaustividad** responde a la fórmula  $a/(a+c)$ , y debería situarse entre el 0,6 y el 0,8.

La tasa de precisión responde a la fórmula  $a/(a+b)$ , y se sitúa entre 0,2 y 0,8.

El principal problema para calcular la tasa de exhaustividad es conocer la variable  $c$ , casi imposible de ajustar en situaciones normales de recuperación de información, por lo que se suelen utilizar técnicas estadísticas y de muestreo para obtener valores fiables.

La evaluación de los resultados es la que determina qué tan efectiva y eficientemente se llevó a cabo el proceso de solución del problema de información. Los puntos más significativos que se deben evaluar son los siguientes:

- ¿El problema de información quedó resuelto?
- ¿Se obtuvo la información que se necesitaba?
- ¿Se tomó la decisión?
- ¿El producto obtenido cumple con los requisitos originalmente establecidos?
- ¿Qué criterios se puede utilizar para hacer evaluaciones?

Otros aspectos que se deben tener en cuenta para evaluar la efectividad del proceso de solución de los problemas, incluyen la cantidad de tiempo empleado en actividades útiles y si hubo algún error de cálculo en la cantidad de tiempo requerido para finalizar las tareas. Esta auto evaluación nos permite mejorar la habilidad futura para resolver con mayor facilidad otros problemas de información.

<http://intranet.logiconline.org.ve/Techinfo/relevancia.html>

## **LA IMPORTANCIA DE LA "RELEVANCIA" EN INFORMACION**

Alfons Cornella.

Leo en el Bulletin de la ASIS (American Society for Information Science) el discurso del brillante profesor Dagobert Soergel con ocasión de haber recibido el Premio de Honor de la ASIS. El discurso, titulado "An Information Science Manifesto" esta accesible en texto entero en <http://www.asis.org/Bulletin/Dec-97/Soergel.htm>. Para cualquiera que quiera enterarse de hacia dónde va esta disciplina, este es un fantástico Manifiesto.

Soergel considera que la tensión entre las necesidades de información del usuario y las formas de representación de esta información en algún tipo de almacén es la clave de la disciplina de la ciencia de la información. En otras palabras, el problema fundamental para quien estudie científicamente el campo de la información es cómo una necesidad de información de alguien es resuelta o no mediante una búsqueda de información en algún sistema que la almacene. Esta relación entre "necesidad de información" e "información almacenada" define el concepto clave: "relevancia".

La idea de relevancia es muy compleja. Primero, y para simplificar, se pueden definir dos tipos fundamentales de relevancia. La relevancia "formal": cuando los resultados de una búsqueda de información responden a la ecuación de búsqueda que se había planteado. Y la relevancia semántica: cuando los resultados obtenidos responden a las necesidades del usuario (he sacado estas dos ideas de un útil librito de J. R. Pérez Alvarez-Ossorio, Introducción a la información y documentación científica).

De estas dos definiciones se entiende ya algo que va a tomar más y más importancia en esta era de Internet: los internautas van a ir descubriendo, ya lo están haciendo de hecho, por qué lo que obtienen en sus búsquedas en el Web no les sirve de mucho. La razón por la que no tan sólo consiguen información muy poco relevante formalmente (sus ecuaciones de búsqueda son simplemente pobres, o el lugar en el que buscan no es el adecuado), sino que lo que consiguen es, frecuentemente, también irrelevante semánticamente, es decir, que no les sirve para resolver la necesidad de información que tenían, porque confunden la información que desean obtener con lo que realmente preguntan. Los grandes depósitos de información que constituyen Internet no sirven de nada si no se les interroga adecuadamente, pero tampoco sirven de nada si quien pregunta no sabe realmente cuál es el problema que quiere resolver buscando información. Por lo general, se dirigen al Web antes de haber meditado sobre cuál es en realidad el problema, es decir, sobre cuál es la información que necesitan. E incluso en ese caso, no se conocen adecuadamente las formas de interrogar el Web. ¿Cuántos internautas, por ejemplo, se han atrevido a indagar cómo se plantean búsquedas

complejas (mediante operadores booleanos, por ejemplo) en los motores de búsqueda universalmente conocidos? O sea: que la práctica en el Web lleva a miles de personas a encontrarse con el problema de la relevancia. Y frente a la irrelevancia de muchas de sus búsquedas sólo se les ocurre dar las culpas a la poca organización de la información en Internet, a la poca calidad de la misma, etc. Una gran oportunidad, de hecho, para que aparezcan intermediarios que prometan una mayor relevancia en la búsqueda de información para profesionales. Algo que, sin embargo, resulta siempre difícil de conseguir.

La principal razón es que la verdadera relevancia, es decir obtener información relevante para un problema, no consiste simplemente en definir una ecuación de búsqueda correctamente, o en ir a hurgar en una buena base de datos. Porque, como nos recuerda Stéfano Mizzaro en un brillante artículo en el Journal del ASIS de septiembre de 1997 (p810), la relevancia es en realidad una relación entre dos entidades cualesquiera escogidas cada una de ellas entre los componentes de dos grupos distintos. En el primer grupo están: el problema (aquel que un humano tiene que resolver), la necesidad de información (cómo el humano en cuestión representa en su mente el problema al que se enfrenta), la petición de información (cómo expresa el humano esa necesidad de información a alguien, normalmente en lenguaje natural), y finalmente, la interrogación (la ecuación de búsqueda a plantear a una base de datos, por ejemplo).

En el segundo grupo están: el documento (el soporte físico donde está la información, por ejemplo un artículo de revista en papel), el subrogado (la representación de ese documento, por ejemplo, el resumen de un documento en una base de datos), y finalmente, la información (lo que el usuario capta al leer un documento).

Así, hablaríamos de diversos tipos de relevancia. Por ejemplo, podemos hablar de un subrogado que es relevante de acuerdo con una cierta interrogación (los registros que se han obtenido de una búsqueda en una base de datos satisfacen correctamente la ecuación de búsqueda planteada).

O podemos hablar de la relevancia de un documento con respecto a una petición (el documento obtenido responde a la necesidad expresada por el peticionario). Pero ninguna de las dos relevancias garantiza que el documento en cuestión responda al problema real del usuario, que, por otra parte, puede que incluso no conozca bien.

Todo ello nos lleva a que va a ser más y más importante instruir a la gente, en especial a los miembros de una empresa, en cómo reconocer bien los problemas, cómo plantearlos en forma de necesidad de información, y cómo plantear ecuaciones de búsqueda que lleven a documentos "relevantes" para el problema, eso dejando de lado la necesidad de conocer bien cuáles son las fuentes de información y cómo se usan. Estas habilidades forman parte de lo que en otras ocasiones hemos denominado "cultura de la información" de los miembros de una empresa.

Este problema, muy conocido por los profesionales de la información desde hace décadas, toma ahora una nueva dimensión con la divulgación universal de Internet.

Publicado en Extra!-Net, edición del 19 de enero de 1998.

## Capítulo 5. Las herramientas de búsqueda de información en Internet

La gran cantidad de información que está almacenada anárquicamente en Internet, generó la necesidad de contar con herramientas llamadas  **motores de búsqueda o buscadores**, que son conjuntos de programas instalados en un servidor conectado a Internet. La búsqueda de información en estos programas puede realizarse de forma temática, a través de índices y subíndices con información previamente indexada, o por palabras. Además existen páginas que funcionan como vínculos para acceder a otros sitios sobre temas determinados. En la búsqueda temática, el programa ofrece al usuario una serie de temas o áreas de interés. Una vez seleccionado uno de ellos, se abren nuevas opciones de subtemas, y así, en pasos sucesivos, se va refinando la búsqueda hasta llegar a la información requerida.

Las búsquedas por palabras clave se activan a través de una o más palabras escritas en la línea correspondiente y el buscador ofrece una página con la lista de sitios que contienen los términos solicitados. Aunque este método es más rápido que el primero, generalmente brindará un resultado sin contexto, trayendo información donde el término requerido puede aparecer empleado en otro sentido.

**Hernández, Tony. Universidad Carlos III de Madrid.**  
**Almacenamiento, búsqueda y recuperación de la información**

<http://www.bib.uc3m.es/~tony/pdmo/pdmotema09.htm>

**Concepto, proceso y sistemas de recuperación de información**

Concepto: Proceso en el que un conjunto de documentos son "interrogados" para encontrar ítems que puedan ayudar a satisfacer las necesidades de información de un usuario.

\* Comercialmente, la recuperación de documentos se ha tratado como una variante de la recuperación de datos.

<b>Recuperación de datos</b>	<b>Recuperación de documentos</b>
<ul style="list-style-type: none"><li>- <u>Recuperación directa</u> que responde la pregunta del usuario. Ej.: Quiero saber X.</li></ul> <p>(Método de respuesta al usuario.)</p>	<ul style="list-style-type: none"><li>- <u>Recuperación indirecta</u> que proporciona o referencia a un conjunto de documentos que probablemente contenga la respuesta a lo que pregunta el usuario. Ej.: Quiero saber sobre X.</li></ul>
<ul style="list-style-type: none"><li>- <u>Necesariamente</u>, existe una relación entre una pregunta bien construida y la respuesta correcta.</li></ul>	<ul style="list-style-type: none"><li>- <u>Probablemente</u>, existe una relación entre una pregunta bien construida y la respuesta correcta.</li></ul>
<ul style="list-style-type: none"><li>- El criterio para medir el éxito de una búsqueda es la <u>corrección</u> de la respuesta. Evaluación objetiva. ¿Es la respuesta correcta?</li></ul>	<ul style="list-style-type: none"><li>- El criterio para medir el éxito de una búsqueda es su <u>utilidad</u>. Evaluación subjetiva. ¿Se ha satisfecho la necesidad del usuario?</li></ul>
<ul style="list-style-type: none"><li>- La velocidad de la recuperación depende sobre todo de la velocidad del <u>acceso físico</u> del sistema que se utilice.</li></ul>	<ul style="list-style-type: none"><li>- La velocidad de la recuperación depende sobre todo del número de <u>decisiones lógicas</u> que el usuario debe tomar durante el desarrollo de la búsqueda y sólo secundariamente de la velocidad del acceso del sistema que se utilice.</li></ul>

## **Directorios y motores de búsqueda**

**Directorios**

**Motores de búsqueda**

un sitio Web. Desde ahí se puede explorar hasta encontrar lo que se busca.	página exacta en las que aparecen las palabras o la frase que uno busca.
Utilice un directorio cuando tenga una vaga idea de lo que quiere y cuando apreciaría que alguien le ayudase.	Utilice un motor de búsqueda cuando quiera llegar a una pieza particular de información muy rápidamente.
Cuando quiera encontrar un buen sitio de música o un sitio especializado en el tipo de películas que a usted le guste, use un directorio.	Cuando quiera saber de qué disco es una canción o una frase de película, utilice un motor de búsqueda.
Utilice un directorio para encontrar las cosas que uno buscaría en las páginas amarillas. Encontrar fontaneros cuando no sabemos el nombre de la empresa de fontanería, etc.	Utilice un motor de búsqueda para encontrar información de un producto en particular cuando sabe el nombre del producto, aunque no su fabricante.
Utilice un directorio cuando quiera navegar, curiosear, cuando le apetezca ver qué hay en la red pero sin buscar nada en concreto.	Utilice el motor de búsqueda cuando tiene poco tiempo o quiere encontrar lo que busca.
Use un directorio cuando busque un sitio dedicado a un personaje célebre.	Use el motor de búsqueda cuando busque a una persona por su nombre.
Use un directorio para encontrar una lista de alumnos o exalumnos de una universidad.	Use un motor de búsqueda para encontrar un artículo concreto escrito por un profesor de antropología de una universidad.
Utilice un directorio para conocer los sitios que estén especializados en buscar trabajos.	Utilice un motor de búsqueda para encontrar el candidato que busca para un trabajo escribiendo las palabras de lo que le gustaría ver en su currículum.

Fuente: Es una traducción libre de un tutorial de búsqueda existente en Altavista. Estas tablas se encuentran accesibles en:

[http://doc.altavista.com/adv\\_search/ast\\_dse\\_uses.shtml](http://doc.altavista.com/adv_search/ast_dse_uses.shtml) (Última visita: 13-03-2000)

#### Bibliografía complementaria

- Altavista: [http://doc.altavista.com/adv\\_search/ast\\_dse\\_uses.shtml](http://doc.altavista.com/adv_search/ast_dse_uses.shtml) (Última visita: 13-

03-2000)

- **Hernández Pérez, Antonio.** "La búsqueda y recuperación de información en Internet." En: La sociedad de la información: Política, Tecnología e Industria de los contenidos. Coordinadora Mercedes Caridad Sebastián. Madrid: Centro de Estudios Ramón Areces, 1999. pp. 213-42.
- **Rodríguez Yunta, L.** "Evaluación e indicadores de calidad en bases de datos." Revista Española de Documentación Científica 21.1 (1998): pp.9-23.

**Es un curso de José Tramullas para el Instituto de Ciencias de la Educación de la Universidad de Zaragoza**

## **Aproximaciones a la recuperación de información**

Los fundamentos sobre los que construir una técnica exitosa de recuperación de información en Internet son el conocimiento de las características propias de los documentos existentes en Internet, y de la teoría de la recuperación de información. Las herramientas de búsqueda aplican sobre el texto los principios que se han explicado sobre recuperación de información textual, y los usuarios disponen de las mismas prestaciones para la recuperación: operadores booleanos, de posición, sectorización, etc. Al tratarse de un entorno abierto y cambiante, las herramientas de búsqueda ofrecen listados de resultados, que dirigen al usuario hacia el documento original. Los cambios que se producen, por la propia dinámica del web, hacen que en ocasiones esa redirección no ofrezca los resultados esperados, y que en numerosas ocasiones haya que completar la búsqueda mediante procesos de exploración basados en la navegación. El usuario siempre debe pensar que no es suficiente, en recuperación de información en Internet, con seguir los resultados obtenidos de un motor de búsqueda: hay que explorarlos, analizarlos, valorarlos, y seleccionarlos como adecuados, o desecharlos como no pertinentes. Las herramientas de recuperación de información en el web son un medio más, una fase intermedia, no un fin.

Las aproximaciones a la RI en Internet pueden establecerse según la creación de recursos, o según la herramienta utilizada. Ambos enfoques, al igual que en casos anteriores, no son excluyentes:

*Según la creación del recurso:*

- La creación de listados, índices y catálogos ordenados por áreas o materias, de forma que el usuario dispusiese de un conjunto de fuentes seleccionadas en las que empezar a buscar. El ejemplo más conocido es el norteamericano Yahoo! Han ido añadiendo motores internos a sus prestaciones, de forma que permiten consultar mediante ecuaciones sus bases de datos
- La creación automática de bases de datos basadas en índices o ficheros inversos, mediante unas aplicaciones que rastrean o exploran todo el ámbito Web, llamados robots, spiders o wanderers. Estos robots rastrean el web a la búsqueda de documentos, obtienen una copia, la indizan según los métodos vistos en el capítulo anterior, y usan los enlaces presentes en los mismos para localizar nuevos documentos

*Según la herramienta utilizada:*

1. Directa: El navegador se conecta al servidor web que actúa como interfaz del motor de búsqueda correspondiente a la base de datos que desea consultar. El servidor le envía una página web que actúa como interfaz de interrogación, a través de la cual formula la consulta. El servidor la recibe, procesa y envía como respuesta una nueva página web, generada de forma dinámica, que contiene las respuestas más pertinentes a la cuestión formulada por el usuario.
2. Por intermediario: El navegador se conecta a un servidor web que le ofrece una interfaz de interrogación propia. Esta interfaz le permite interrogar una base de datos correspondiente a un motor de búsqueda situado en un servidor web diferente al que ofrece la interfaz. La interfaz actúa como intermediario entre el motor de búsqueda de destino y el usuario. El interés de estos intermediarios (metabuscadores) se da cuando consultan a múltiples motores de búsqueda.
3. Por agente: El usuario instala en su máquina una aplicación que permite formular las ecuaciones de búsqueda y remitirlas a uno o varios motores de búsqueda. La aplicación lanza conexiones simultáneas al conjunto de motores que se trate, recibe las respuestas, y las entrega al usuario en una presentación única, que puede ofrecer diferentes formas. Dependiendo de las prestaciones del agente, las respuestas pueden ser filtradas, aplicando criterios propios de eliminación de duplicados, reordenación de resultados, etc.
4. Por robot personal: Se trata de aplicaciones que se instalan en el ordenador del usuario, y que son capaces de acceder a un servidor web, construir un mapa de índices de sus contenidos, y utilizar los mismos para acceder a la información que sea interesante para el usuario, obteniendo copias de las páginas o documentos web contenidos en el mismo.

# **La Navegación Hipertextual en el World Wide Web: Implicaciones para el Diseño de Materiales Educativos**

**Jordi Adell**

## **1. El World-Wide Web**

El World Wide Web (WWW) es un sistema hipermedia distribuido, accesible a través de Internet, que permite navegar con facilidad por una enorme cantidad de información. El WWW fue iniciado en el CERN por Tim Berners-Lee (Berners-Lee, 1993a; Berners-Lee, Caillau, Groff y Pollerman, 1992a y 1992b) con el objeto de integrar información accesible a través de una única red de ordenadores, pero mediante sistemas diversos.

El WWW se sustenta en cuatro elementos fundamentales:

1. Un nuevo protocolo de comunicación (HTTP o HyperText Tranfer Protocol);
2. Un lenguaje para escribir documentos hipermedia (HTML o Hypertext Markup Language);
3. Un sistema notacional para designar objetos en la Internet y las operaciones a realizar sobre ellos (URL o Uniform Resource Locator); y, finalmente,
4. Un conjunto de aplicaciones (los clientes o browsers WWW y los servidores tgh) que se dividen el trabajo de servir y presentar la información multimedia al usuario.

Los elementos básicos de la tecnología WWW se han descrito ya en diversas publicaciones en nuestro país (Adell, 1993; Adell y Bellver, 1993 y 1995). El número de servidores WWW instalados crece a un ritmo vertiginoso (en nuestro país superan los 300) y se estima que, a nivel mundial, su número se dobla cada 50 días. La cantidad de información accesible mediante el WWW está en el orden de terabytes. Su uso masivo se realiza en campos como la comunicación científica, los negocios, el ocio y la educación.

## **2. El WWW en educación**

La utilidad educativa del WWW es evidente. No sólo participa de las características de un sistema hipermedia tradicional, sino que puede utilizarse ventajosamente en campos como la educación a distancia o la elaboración de materiales de enseñanza/aprendizaje interactivos y/o permanentemente actualizados. Un estudiante, desde su casa o desde un puesto de trabajo en su centro de estudios, puede acceder con rapidez a grandes cantidades de información, estructurada con criterios didácticos, que puede residir físicamente en uno o varios ordenadores locales o remotos.

Las características que hacen del WWW una tecnología de amplio potencial educativo son, a nuestro juicio, las siguientes:

- a) Capacidad hipertexto/hipermedia: la estructura de la información no es lineal, sino hiperdimensional. Es posible diseñar materiales adaptados a diferentes niveles, expectativas, etc. de los aprendices y estructurar la información de modo de los lectores construyan sus propios significados seleccionando qué nodo o lexia examinarán y cual soslayarán.
- b) Capacidad multimedia: mediante el WWW pueden distribuirse documentos multimedia (texto, imágenes, fragmentos de video, animación, sonido, aplicaciones informáticas, consultas online a bases de datos, formularios, mapas sensibles, etc.).
- c) Capacidad como sistema distribuido y abierto a la Internet: mediante el WWW es posible la construcción de hipermedia complejos almacenados en diferentes servidores de la Internet y, por tanto, el trabajo colaborativo entre equipos de investigadores y profesores. El acceso desde cualquier ordenador conectado a la Internet permite su utilización como sistema de educación electrónica a distancia, como "aula virtual" de enseñanza/aprendizaje en la que los estudiantes y sus profesores se comunican en tiempo real o diferido mediante diversas aplicaciones de comunicaciones (video conferencia, talk, correo electrónico, pizarras electrónicas, etc.).
- d) La disponibilidad gratuita de clientes, servidores, aplicaciones auxiliares para la visualización y audición de formatos diversos (texto, gráficos, audio, video, sesiones interactivas, pasarelas a otros sistemas, etc.) y para la comunicación, herramientas para la elaboración de hipermedia y de gestión de servidores, etc. para cada casi cualquier tipo de plataforma hardware/software pone la tecnología WWW al alcance de cualquier persona o grupo con acceso a la Internet.
- e) Capacidad interactiva ampliada: los formularios y scripts CGI permiten que el usuario interactúe con el sistema de modo más completo que mediante la navegación por la información.

Respecto al uso educativo del WWW, Ibrahim y Franklin (1995) han identificado dos ejes fundamentales: como sistema hipermedia cerrado y como sistema hipermedia abierto a la Internet. Un sistema hipermedia cerrado se caracterizaría por limitar todos los vínculos contenidos en las páginas que componen el documento hipermedia a nodos controlados, es decir, a evitar la navegación azarosa y a circunscribir las posibilidades del aprendiz a un conjunto finito y cerrado de nodos. El WWW permitiría distribuir courseware a través de Internet a localizaciones remotas, pero este tipo de material sería similar al que puede realizarse con los sistemas hipermedia en los que toda la información reside en el ordenador local. En esta línea se están realizando diversas experiencias, demasiado numerosas para citar aquí más que algunas muestras. Por ejemplo, en Campbell, Hurley, Jones y Stephens (1995) se describen los pormenores de un

proyecto de construcción de courseware sobre supercomputación. Dwyer, Barbieri y Doerr (1995), del Cornell Theory Center, han diseñado el tipo de "aula virtual" necesaria para desarrollar este tipo de enseñanza/aprendizaje. Bilotta, Fiorito, Iovane y Pantano (1995) han construido un "ambiente educativo" electrónico centrado en el estudiante mediante el WWW en el Centro Interdipartimentale della Comunicazione de la Universidad de Calabria (Italia).

Un proyecto de coordinación de esfuerzos en la elaboración y distribución de materiales educativos mediante el WWW es la Global Network Academy (Butts, Reilly, Speh y Wang 1994), en la que pueden encontrarse numerosas muestras de material diseñado para la enseñanza en el marco del WWW.

El otro enfoque propuesto por Ibrahim y Franklin (1995) intenta explotar el enorme potencial de la Internet, es decir, la vasta cantidad de información, datos, documentos, imágenes, etc. accesible mediante el WWW, integrándola de manera dinámica en el material educativo. Para integrar recursos útiles de entre esta maraña de información disponible es necesario contar con catálogos (como la Biblioteca Virtual del CERN), revistas electrónicas, colecciones de recursos categorizados por temas (ciertos FAQs, por ejemplo) y, en general, herramientas de búsqueda y recuperación de información poderosas y sencillas de utilizar. Los materiales desarrollados desde este enfoque podrían beneficiarse de las ventajas de la actualización automática de información, creándose "documentos dinámicos" (Adell, 1993), y de la labor desinteresada de numerosos expertos que mantienen colecciones temáticas de punteros a multitud de recursos de la red. Evidentemente, este tipo de materiales tiene su público natural entre los estudiantes universitarios o de doctorado. En niveles más bajos es necesario "controlar" los materiales para asegurar la adecuada asimilación de contenidos mínimos.

Un documento hipermedia desarrollado desde esta perspectiva abierta sería un conjunto de explicaciones (texto y gráficos), que incluirían punteros a materiales de mayor especificidad y profundidad disponibles en la Internet o a información en tiempo real, e incluso algún tipo de formulario de autoevaluación que, una vez puntuado automáticamente, remitiría al estudiante a ciertos materiales en función de su rendimiento en las distintas partes de la unidad didáctica. Sin embargo este enfoque precisa estudiantes avezados en el uso del WWW y la Internet y herramientas poderosas de navegación (y, desde luego, cierta autodisciplina) a fin de no perderse en el hiperespacio de la información.

El uso educativo del WWW presenta, no obstante sus potencialidades, problemas diversos. A los inherentes al diseño de hipermedia con finalidad educativa, hay que añadir la integración de materiales dinámicos y cambiantes, como los que caracterizan la Internet o los derivados del crecimiento explosivo de los usuarios y del tráfico de la red. La lentitud con la que se recupera la información en países en los que la infraestructura de la Internet no es suficiente limita el uso de elementos multimedia a simples "demos" o decoración. La información audiovisual digitalizada precisa líneas de alta capacidad. Sin embargo, la cuestión en la que

vamos a centrarnos en las siguientes secciones hace referencia al diseño de materiales educativos y a evitar un fenómeno recurrente en la literatura hipertextual y que todo usuario del WWW ha sentido alguna vez: los usuarios "perdidos en el hiperespacio".

### **3. Perdidos en el hiperespacio**

*"Cuando los usuarios se mueven por un espacio amplio de información como en el caso de hipertextos, existe un riesgo real de que se desorienten o tengan problemas para encontrar la información que necesitan" (Nielsen, 1990).*

Diseñar hipermedia es tanto una cuestión de contenidos como de relaciones entre unidades de contenido o lexias (cómo unirlas formando una estructura hipertextual coherente y utilizable para el aprendizaje). Los problemas derivados de la navegación por la información han sido un tema recurrente en la literatura sobre hipertexto (Edwards y Hardman, 1989). Se ha acuñado una expresión para definir el problema típico de la navegación: "perdidos en el hiperespacio". Este tipo de síndrome se produce cuando el lector:

- No sabe dónde se encuentra ahora.
- No sabe cómo volver a algún lugar conocido.
- No sabe cómo buscar la información que necesita o desea.
- Tiene la sensación de que, a pesar de sus esfuerzos, se está perdiendo algo importante.

El WWW en tanto que sistema hipermedia distribuido posee características que "facilitan" el que los usuarios se "pierdan en el hiperespacio". Entre ellas podemos citar las siguientes:

1. Un usuario puede llegar a una página de un servidor WWW público desde cualquier nodo de la Internet. El botón de "vuelta atrás" de su browser le retrotrae a la página que ha visitado antes, no a la que lógicamente antecede a la actual. Es más, a una página o documento dado, los usuarios pueden haber llegado por diferentes caminos. No debe asumirse que han visto/leído todos lo mismo.
2. La flexibilidad del WWW permite establecer vínculos en un nodo con casi cualquier objeto de la Internet recuperable mediante casi con cualquier protocolo (o a través de pasarelas), tiene su contrapartida en usuarios "perdidos": ofrecemos un "vínculo" a un documento almacenado en otro servidor de la red y "nuestros" usuarios se pierden en ese otro servidor o bien el objeto recuperado es un documento de texto sin vínculos a otros documentos ni dispositivos de navegación: un "callejón sin salida".
3. La utilización de aplicaciones auxiliares para acceder a la información recuperada en formatos que los browsers no pueden manejar (el sonido, los fragmentos de video o la imagen de calidad (JPEG), por ejemplo) puede desorientar a usuarios poco duchos: "salen" del browser y la nueva aplicación no

tiene nada parecido a un botón que diga "volver al cliente WWW que estaba utilizando hace un momento".

4. La, cada vez mayor, cantidad de documentos HTML generados on the fly por scripts CGI, que no suelen incorporar dispositivos de navegación. Suelen ser "callejones sin salida" en los que sólo queda "hacer marcha atrás" para volver a algún lugar conocido (generalmente el formulario en respuesta al cual se ha generado el documento HTML actual).

5. Las páginas de menús, a las que tan aficionados son los autores de páginas HTML, pocas veces incluyen en cada alternativa información suficiente para que el usuario decida con conocimiento de causa qué camino tomar. Denominaciones crípticas, menús oscuros, alternativas disimilares, siglas abstrusas, etc. son la causa de que los lectores se vean sorprendidos en ocasiones con contenidos escasamente relacionados con sus expectativas

6. Los usuarios de un servidor WWW tienen diferentes intereses, nivel de formación, expectativas, experiencia con el sistema, etc. Con usuarios perfectamente identificados y conocidos (una clase, un grupo de interés, etc.) es posible afinar mucho el diseño de un hipermedia.

7. Finalmente una consideración pedagógica: ¿cómo garantizar que los lectores de un hipermedia cuyo objeto es formativo recorran los nodos principales y no se pierdan por la miríada posible de relaciones hipertextuales? Evidentemente, limitando de alguna manera los caminos no sustantivos en el diseño del hipertexto y ofreciendo herramientas para la navegación.

Por tanto, en los sistemas hipermedia y especialmente en el WWW, que permite incluir como nodo en nuestro hipermedia casi cualquier objeto de la Internet, la navegación por la información es un problema real.

Este tipo de problemas se han intentado solucionar con diversas medidas. A nuestro juicio, las claves en el diseño de la navegación de un servidor WWW están en:

1. Un diseño claro de la estructura del hipermedia, de la que se debe hacer consciente al usuario (incluyendo meta-information, por ejemplo).

2. Una elaboración cuidadosa de las páginas que tenga en cuenta los diferentes tipos de usuarios potenciales y que incluya la suficiente información para que el lector pueda decidir entre las diversas opciones que se le ofrecen.

3. Un conjunto de dispositivos de navegación que permitan a los lectores:

- determinar su localización actual

- hacerse una idea de la relación de tal localización con otros materiales
- volver al punto de partida, y
- explorar materiales no directamente vinculados a los que se encuentra en este momento (*Landow, 1992, pág. 86*).

#### 4. Herramientas para la navegación hiperespacial

Navegar por el hiperespacio de la información es un problema, sobre todo en sistemas grandes. Diversos autores han intentado descomponer el problema. Así, por ejemplo, Wright y Lickorish (1990), y **en relación al sistema**, han distinguido entre navegación **interna**, esto es, aquella que forma parte del hipertexto, y **externa**, aquella posibilitada por las herramientas de navegación genéricas, independientes del hipertexto. En sus experimentos los lectores preferían el uso de un índice externo (frente a links internos) en un material hipertextual similar a un libro. En otro hipertexto, menos parecido a un libro, los lectores prefirieron la navegación interna. Su conclusión es que combinar sistemas de navegación puede ampliar el rango de tareas para las que pueden utilizarse hipertextos.

Monk (1990), analizando las **pautas de conducta de los usuarios**, ha distinguido entre navegación **exploratoria y dirigida**. Es dirigida cuando el usuario ha formulado una sub-meta para acceder a una localización conocida dentro del hipertexto. La navegación exploratoria, en cambio, no tiene destino previamente determinado: el usuario se guía por sus intereses y lo que le sugiere la información asociada a los vínculos. Estos usuarios "callejean" por el hipertexto. Puede que al final se sientan perdidos si intentan regresar a algún lugar determinado, pero su objetivo es examinar el contenido del hipertexto sin un plan sistemático. Catledge y Pitkow (1995), analizando los registros de actividad de un servidor y diversos clientes en el Georgia Institute of Technology, han caracterizado la conducta de los usuarios en el WWW en tres categorías:

- a) *Search browsing*: búsqueda directa con una meta final conocida.
- b) *General purpose browsing*: consulta de fuentes que tienen una alta probabilidad de contener ítems de interés.
- c) *Serendipitous browsing*: puramente al azar, guiados por el atractivo de los ítems.

La conducta de cada tipo de usuario determina qué facilidades de navegación espera y usará y qué tipo de decisiones de diseño es necesario adoptar.

Las siguientes herramientas de navegación representan soluciones acuñadas con el tiempo y la experiencia acumulada con diversos sistemas hipertexto por numerosos autores. Algunos están directamente implementados en los browsers WWW, pero la mayoría quedan a la discreción del autor del hipertexto.

#### 4.1. La "vuelta atrás" y la historia de nodos visitados

La **vuelta atrás** es el mecanismo que permite que un usuario o lector regrese sobre sus propios pasos nodo a nodo (i.e., botón de retroceso) o saltando a cualquier nodo visitado anteriormente (**historia de nodos visitados**). Evidentemente el camino recorrido hacia atrás debe poderse recorrerse nuevamente hacia adelante.

La vuelta atrás permite que los usuarios perdidos regresen a jalones significativos de su camino (véase más abajo) o que exploren otros nodos que, en la primera pasada, no seleccionaron. Las "encrucijadas" son la situación normal en un hipertexto rico en vínculos. En un "cruce de caminos", un usuario escoge uno de ellos. Eso no implica que los demás no le interesan: debe dársele la oportunidad de que regrese y reconstruya su camino por otros senderos.

Todos los browsers WWW incluyen sistemas (activados mediante botones o comandos) de "atrás", "adelante" y "home" o página inicial. Esta facilidad puede entrar en contradicción con el esquema lógico de lectura de hipertextos cuando los usuarios "llegan" a mitad de un hipertexto procedentes de otro servidor de la red. Por eso es necesario implementar a nivel de documento herramientas de navegación en forma de botones o links que definan la navegación deseada por el autor del hipertexto y exigida por la naturaleza del documento.

#### 4.2. Visitas guiadas

Tal vez la solución más simple al problema de "perderse en el hiperespacio" desde la perspectiva del lector sea eliminar la necesidad de "navegar". Las **visitas guiadas** (Trigg, 1988; Nielsen, 1990, pág. 127 y ss.) son la forma de acompañar a los usuarios por un camino prefijado, de llevarles de la mano durante sus primeras experiencias con el hipertexto, sin eliminar las posibilidades de navegación. La visita guiada es útil para introducir a los nuevos usuarios en los conceptos generales del hipertexto (Nielsen, 1990, pág. 128) o en las características distintivas de un hipertexto en particular.

El concepto de camino o sendero entre nodos de información se remonta al propio Bush (1945). El "memex", el dispositivo esbozado por Bush, debía almacenar enormes cantidades de información. El usuario podría crear "pistas" o "senderos", es decir, conexiones entre textos e ilustraciones relacionados. Estas "pistas" podrían guardarse y servir para futuras exploraciones.

Una visita guiada puede definirse como una secuencia de vínculos entre nodos, una "cadena" de nodos, que se ofrecen al usuario para una lectura secuencial y que proporcionan una visión determinada de un hipertexto. Un hipertexto puede tener varias "visitas guiadas", cada una respondiendo a necesidades diversas de los usuarios (por ejemplo, una visita puede diseñarse para que los usuarios neófitos se hagan una idea general sobre qué van a encontrar, otras pueden resolver necesidades informativas típicas de usuarios determinados, etc.).

La existencia de visitas guiadas no supone la desaparición de los otros vínculos entre nodos, simplemente se destacan o privilegian unos cuantos para cumplir determinado propósito (Marshall y Irish, 1989). El usuario siempre puede "dejar" la visita guiada y seguir su propio camino. Un sistema ideal debería permitir al usuario abandonar la visita guiada en un momento determinado y volver a dicho punto cuando lo deseé.

La visita guiada, sin embargo, no es la solución de todos los problemas de navegación, consustanciales al concepto de hipertexto como "espacio exploratorio y abierto de información" (Nielsen, 1990, pág. 128). La visita guiada, en cierta forma, elimina la dimensión hiperespacial retrotrayéndonos a la lectura secuencial.

Nicol, Smeaton y Slater (1995) han desarrollado un sistema de visita guiada para el WWW denominado Footsteps. Un script CGI controla las páginas que el usuario recupera desde que comienza la visita guiada, incluyendo links al siguiente paso de dicha visita o, si el usuario la abandona, un botón para retornar al punto en que la abandonó y seguir la visita guiada. El precio de este sistema es que todas las páginas deben ser preprocesadas por el script en tiempo real antes de que el usuario las vea, lo cual incrementa el tráfico de la red y hace más lenta la recuperación. Sin embargo, instalado localmente, el sistema Footsteps permite elaborar hipermedia didácticos y ofrecer a los usuarios la posibilidad de volver a la secuencia principal en caso de que se pierda.

#### **4.3. Diagramas y mapas del espacio hipertextual**

Los diagramas o mapas parecen recursos adecuados para facilitar la "navegación" por la información. Se trata de una representación esquemática del espacio en el que se mueve el lector, incluso señalándole dónde se encuentra en ese momento y la posibilidad de saltar directamente a otras secciones (mapas sensibles con espacios definidos visualmente como vínculos). Los diagramas o mapas que incluyen sólo grupos de nodos fuertemente relacionados simplifican los mapas (evitando la sobrecarga cognitiva) y facilitan percepciones globales de la estructura general. Si cada grupo o cluster posee nodos-índice o jalones (véase el apartado sobre este aspecto), la navegación se facilita al simplificar la estructura.

La perspectiva "ojo de pez" en los mapas, que destaca los espacios adyacentes al que se encuentra el lector, frente a los espacios "lejanos", conceptual e hipertextualmente hablando, muestra en un único gráfico todo el espacio de información pero con niveles variados de detalle. El "ojo de pez", sin embargo, como señala Nielsen (1990) requiere que pueda estimarse la distancia entre una localización determinada y el objeto de interés del usuario y, además, debe ser posible mostrar la información a diversos niveles de detalle. Estas dos condiciones sólo las reúnen hipertextos estructurados jerárquicamente. En otro tipo de estructuras es imposible de utilizar.

Los mapas, sin embargo, tienen sus propias dificultades: son costosos de diseñar manualmente si el espacio de la información es grande. De hecho hay un

momento en el que dejan de ser aclarativos: si pretendemos incluir todo nodo y todo vínculo. Por su parte, los sistemas automatizados, como WebMap (Dömel, 1994) o el Navigational View Builder (Mukherjea y Foley, 1994 y 1995; Mukherjea, Foley y Hudson, 1995) son todavía excesivamente complejos y no están disponibles para las plataformas habituales en nuestras aulas.

#### **4.4. Landmarks**

Otra técnica es distinguir ciertos nodos respecto a los cuales el usuario pueda situarse y ubicar la información. Una metáfora turística nos ayudará a explicar esta técnica: recién llegados a una ciudad desconocida, ciertos edificios o lugares emblemáticos nos sirven para orientarnos y situar otros lugares. Esta técnica forma parte del bagaje del diseñador: uno o varios lugares "clave" desde los que su pueda contemplar y acceder a todo el espacio informativo y a los que se pueda volver desde cualquier sitio con facilidad (la home page, los índices de cada sección, etc.) son imprescindibles.

Un problema del hipertexto (del ordenador) frente al texto impreso tradicional es su homogeneidad: todas las páginas son iguales. Cuando leemos un libro tenemos constancia de dónde nos encontramos. Cuando leemos un hipertexto es difícil aprehender el contexto. La diferenciación (con títulos y subtítulos, colores, etc.) de las diferentes secciones o partes lógicas de un hipertexto ayuda a encontrar de nuevo el camino (y ser consciente de cuando se deja para entrar en otro).

#### **4.5. Índices**

A lo largo de la evolución de la letra impresa se han creado una serie de convenciones que ayudan al lector a encontrar lo que busca dentro de grandes corpus de información impresa. Los índices son uno de los mecanismos más habituales. En el hipertexto se han utilizado tanto índices alfabéticos como jerárquicos. Los índices alfabéticos permiten que los usuarios encuentren la información que buscan en corpus escasamente estructurados. Los jerárquicos, en cambio, parecen más indicados para documentos estructurados. Simpson y McKnight (1990) encontraron que los sujetos a los que se les facilitaba un índice jerarquizado hacían mapas conceptuales del hipertexto más exactos que aquellos a los que se les proporcionaba un índice alfabético.

Sin embargo, la elaboración de índices puede ser una tarea ardua en hipertextos de gran tamaño. Una solución habitual es la construcción de una base de datos full text (como WAIS (Khale, 1991; Marshall, 1992) y sus evoluciones) interrogable mediante pasarelas desde el mismo sistema hipertextual. El lector puede introducir en lenguaje natural términos relativos a la información que busca y la base de datos le indica aquellos documentos del hipertexto en los que aparecen dichas palabras de forma significativa. También es posible restringir o ampliar las búsquedas mediante operadores booleanos y, en el caso de documentos semiestructurados, limitar la búsqueda a alguno de los campos.

## 5. Conclusiones (o "para diseñar hipermedia no basta con saber HTML")

*"Los autores de hipermedia afrontan problemas similares a los que confrontaron los pioneros del cine. Tienen que inventar el primer plano, el fundido a negro y el dissolve. Y cuando llegó el sonido tuvieron que reinventar el medio para incorporarlo. Lo mismo es cierto para los hipermedia. Tenemos la tecnología, pero estamos aún en el proceso de inventar el lenguaje y las convenciones de este nuevo medio de comunicación (Cotton y Oliver, 1993).*

La navegación por espacios hiperdimensionales de información se convierte en un problema a medida que aumenta su tamaño y complejidad. El uso educativo del hipertexto precisa poner en manos de los usuarios hipermedia bien diseñados y herramientas para la navegación. La explosión de la Internet y del WWW ha provocado la masiva creación de hipermedia en los que, en demasiadas ocasiones, no se refleja el bagaje de conocimientos y know how adquiridos durante años sobre otros sistemas (Intermedia, Guide, incluso HyperCard, etc.). Durante el proceso de diseño de un hipermedia es necesario considerar cuidadosamente no solo las diferentes páginas y elementos que las componen, sino las relaciones que se establecen entre éstas. Las posibilidades de un hipermedia, aquello que lo caracteriza y diferencia de los materiales impresos, residen en cómo se puede navegar por la información: frente a la propuesta única del impreso, el hipertexto ofrece cierto grado de libertad al lector para construir sus propios significados. En educación, con objetivos de aprendizaje definidos, el diseño de la navegación es crucial. Con un sistema adecuado de navegación se pueden elaborar hipermedia que respondan a distintos niveles de conocimientos previos, necesidades, objetivos, etc. de los aprendices.

Todavía nos encontramos en la infancia de este tipo de materiales por lo que es necesario experimentar en condiciones realistas de enseñanza/aprendizaje. Precisamos métodos y herramientas para analizar el comportamiento de los estudiantes ante los materiales y explorar las posibilidades interactivas de los nuevos medios. También, a la vista de lo producido hasta la fecha, parece necesario releer a los clásicos: algunos de los problemas de los que adolecen muchos materiales hipertextuales distribuidos mediante el WWW ya surgieron y fueron solucionados en sistemas anteriores. La Internet y los desarrollos actuales en materia de comunicaciones nos ofrecen muchas posibilidades. El WWW, en tanto que sistema hipermedia distribuido, además, tiene su propia retórica. Los autores tienen la obligación, si quieren comunicarse efectivamente con los lectores, de utilizar las técnicas a su alcance. El medio impone sus propias normas sobre contenido y, del mismo modo que el libro, el periódico, la película o el noticario televisivo tienen sus características distintivas, las "leyes del medio", los materiales diseñados para ser "consumidos" mediante el ordenador a través de la red tienen su propio lenguaje. A nosotros nos corresponde investigar sus potencialidades educativas.

## 6. Referencias

- Adell, J. (1993). World-Wide Web: un sistema hipermedia distribuido para la docencia universitaria. I Congreso sobre Nuevas Tecnologías de la Información y la Comunicación aplicadas a la Educación, Badajoz, 17-21 de diciembre de 1993. Publicado en Blazquez, F., Cabero, J. y Loscertales, F. (Coord.). (1994). Nuevas tecnologías de la Información y la Comunicación para la Educación. Sevilla: Ediciones Alfar, pág. 114-121.
- Adell, J. (1995). Educación en la Internet. Universitas, Serie IV, Vol. Extraordinari XX Setmana Pedagògica (ISSN 0211-3368), pág. 207-214.
- Adell, J. y Bellver, C. (1994). Hipermedia distribuido en el Mac: el proyecto World-Wide Web. Presentación online en UNIMAC: I Jornadas Macintosh y Universidad, Universidad Nacional de Educación a Distancia, Madrid, Septiembre de 1994. Publicado en Actas de las I Jornadas Macintosh y Universidad, Madrid: UNED, 1994. También en el CD ROM UNIMAC 1994. y en el MacUser CD ROM nº 1 (1994).
- Adell, J. y Bellver, C. (1995). La Internet como telaraña: el World-Wide Web. Métodos de Información, 2(3), enero 1995. También en Servicom Magazine, n. 3, Mayo de 1995, pág. 10-18. URL:<http://www.uv.es:80//mei3/Web022.html>
- Berners-Lee, T. (1993a). The World-Wide Web Initiative. URL:<ftp://info.cer.ch/pub/www/doc/inet92.ps>
- Berners-Lee, T., Caillau, R., Groff, J-F. y Pollerman, B. (1992a). World-Wide Web: An Information Infrastructure for High-Energy Physics, Presented al Software Engineering, Artificial Inteligence and Expert Systems for High Energy and Nuclear Physics, at Londe-les-Maures, January 1992.
- Berners-Lee, T., Caillau, R., Groff, J-F. y Pollerman, B. (1992b).World-Wide Web: The Information Universe, Electronic Networking: Research, Applications and Policy, Vol 1, No 2, Meckler, Westport CT, Spring 1992.
- Bernstein, M. (1991). The Navigation Problem Reconsidered, en Berk, E. y Devlin, J. (Eds.). Hypertext/Hypermedia Handbook, McGraw-Hill: New York, págs. 285-297.
- Bilotta, E., Fiorito, M., Iovane, D. y Pantano, P. (1995). An Educational Environment Using WWW. Proceedings of the Third International World-Wide Web Conference, Darmstadt, Germany, April 1995. URL:  
<http://www.cubo20.unical.it/server/ricerca/articoli/eduwww.htm>

Bush, V. (1945). "As We May Think", Atlantic Montly, 176/1, July, pp. 101-108. Puede verse la traducción al castellano de este texto en Lambert, S. y Ropiequet, S. (Eds.) (1987). CD ROM. El nuevo papiro. Madrid: Anaya-Multimedia, pp. 3-21.

Butts, C., Reilly, C., Speh M. y Wang, J. (1994). WWW and the Global Network Academy. En Proceedings of the First International on the World-Wide Web, Paper nº 36, CERN Geneva, Swtlzerland, May 25-27, 1994.  
URL:<http://www1.cern.ch/papersWWW94/speh.ps>

Campbell, J.K., Hurley, S., Jones, S.B. y Stephens, N.M. (1995). Constructing Educational Courseware using NCSA Mosaic and the World-Wide Web. Proceedings of the Third International World-Wide Web Conference, Darmstadt, Germany, April 1995. URL:  
<http://www.igd.fhg.de/www/www95/proceedings/papers/52/www3.html>

Catledge, L.D. y Pitkow, J.E. (1995). Characterizing Browsing Strategies in the World-Wide Web. Proceedings of the Third International World-Wide Web Conference, Darmstadt, Germany, April 1995.  
URL:<http://www.igd.fhg.de/www/www95/proceedings/papers/80/userpatterns/UserPatterns.Paper4.formatted.html>

Cotton, B. y Oliver, R. (1993). Understanding Hypermedia, London: Phaidon Press.

Dwyer, D., Barbieri, K. y Doerr, H.M. (1995). Creating a Virtual Classroom for Interactive Education on the Web. Proceedings of the Third International World-Wide Web Conference, Darmstadt, Germany, April 1995.  
URL:<http://www.igd.fhg.de/www/www95/proceedings/papers/62/ctc.virtual.class/ctc.virtual.class.html>

Dömel, P. (1994). WebMap: A Graphical Hypertext Navigation Tool. Proceedings of the Second International WWW Conference, Chicago, 1994.  
URL:<http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/doemel/www-fall94.html>

Edwards, D.M. and Hardman, L. (1989). 'Lost in Hyperspace': Cognitive Mapping and Navigational in a Hypertext Environment. McAleese, R. (Ed.) Hypertext: Theory into Practice. Intellect Press: Oxford, págs. 104-125.

Gay, G. y Mazur, J. (1991). Navigating in Hipermedia, en Berk, E. y Devlin, J. (Eds.). Hypertext/Hypermedia Handbook, McGraw-Hill: New York, págs. 271-283.

Kahle, B. (1991). Overview of Wide Area Information Servers. Thinking Machines.  
URL:[ftp://ftp.wais.com/\\*](ftp://ftp.wais.com/*)

Landow, G.P. (1992). The Rethoric of Hypermedia: Some Rules for Authors. En Delany, P. and Landow, G.P. (1992). *Hypermedia and Literary Studies*. Cambridge, MA: MIT Press.

Marshall, C.C. y Irish, P.M. (1989). Guided Tours and On-Line Presentations: How Authors Make Existing Hypertext Intelligible for Readers. *Hypertext'89 Proceedings*. Association for Computing Machinery (ACM), November 5-8, Pittsburg, Pennsylvania.

Marshall, P. (1992). WAIS: The Wide Area Information Server or Anonymous What??. The University of Western Ontario. Computing and Communications Services. Documento electrónico. URL:<ftp://ftp.wais.com/pub/wais-doc/UWO-wais-paper.ps>

Monk, A.F. (1990). Getting to know locations in a hypertext. En McAleese, R. y Green, C. (Eds.). *Hypertext: State of the Art*. Oxford, England: Intellect.

Mukherjea, S. y Foley, J. (1994). Navigational View Builder: A Tool for Building Navigational Views of Information Spaces. En ACM SIGCHI'94 Conference Companion, pp. 289-290. Boston, Ma. Abril 1994.

Mukherjea, S. y Foley, J. (1995). Visualizing the World-Wide Web with the Navigational View Builder. *Computer Networks and ISDN System, Special Issue on the Third International Conference on the World-Wide Web '95*, April 1995, Darmstadt, Germany.

URL:<http://www.igd.fhg.de/www/www95/proceedings/papers/44/mukh/mukh.html>

Mukherjea, S., Foley, J y Hudson, S. (1995). Visualizing Complex Hypermedia Networks through Multiple Hierarchical Views, ACM SIGCHI 1995, May 1995, Denver, Colorado.

URL:[http://www.cc.gatech.edu/gvu/people/Phd/sougata/chi95/sm\\_bdy.html](http://www.cc.gatech.edu/gvu/people/Phd/sougata/chi95/sm_bdy.html)

Neuss, C. y Höfling, S. (1995). Lost in Hyperspace? Free Text Searches in the Web, Proceedings of the First International World-Wide Web Conference, pp 121-128 CERN, Switzerland, May 1994. URL:<http://www.igd.fhg.de/~neuss/w4-main.html>

Nicol, D., Smeaton, C. y Slater, A.F. (1995). Footsteps: Trail-blazing the Web. Proceedings of the Third International Conference on the World-Wide Web, April 1995, Darmstadt, Germany.

URL:<http://www.igd.fhg.de/www/www95/proceedings/papers/60/footsteps.html>

Nielsen, J, (1990). The Art of Navigating Through Hypertext, *Communications of the ACM*, 33(3), pp. 296-310.

Nielsen, Jakob (1990). *Hypertext and Hypermedia*. Academic Press: San Diego, CA.

Simpson A. y McKnight, C. (1990). Navigation in hypertext: structural cues and mental maps. En McAleese, R. y Green, C. (Eds.). *Hypertext: State of the Art*. Oxford, England: Intellect.

Trigg, R.H. (1988). Guided tours and tabletops: Tools for communicating in a hypertext environment. *ACM Trans. Office Information Systems* 6(4), págs. 398-414.

Wright, P. y Lickorish, A. (1990). An empirical comparison of two navigation systems for two hypertexts. En En McAleese, R. y Green, C, (Eds.). *Hypertext: State of the Art*. Oxford, England: Intellect.

## **Técnicas de Recuperación de Información**

**Universidad Cardenal Herrera CEU, Valencia - España**

**Julián Gómez Ayora & Daniel Fernández Juez**

**La recuperación de información** es el proceso que permite "obtener de un conjunto de documentos, aquellos que se adecuan más a una demanda de información". Para poder realizar esta recuperación es necesario realizar previamente una búsqueda adecuada.

### **¿Que busco y para que lo busco?**

Una búsqueda es difícilmente fructuosa si no se sabe con claridad lo que se busca, y menos en Internet que, a parte de alojar una enorme cantidad de información, carece de cualquier tipo de estructura que pueda facilitar la búsqueda. También hay que tener en cuenta que la cantidad de información ofrecida no implica que esta sea de calidad o que esté adecuada a las necesidades del usuario. Unas preguntas que es aconsejable plantearse antes de iniciar la búsqueda son:

- ¿Qué tipo de información necesitamos? (Libros, Artículos, Informes, Direcciones)
- ¿Cuáles son los idiomas de preferencia?
- ¿Necesitamos información muy especializada?
- ¿Cuánto tiempo estamos dispuestos a invertir?

### **¿Donde puedo encontrarlo?**

Cada necesidad de información tiene asociada los tipos de fuentes de información más adecuadas. El tipo de información que busquemos nos indicará que fuentes son las más idóneas para contestar nuestras preguntas. No es lo mismo buscar artículos técnicos, un número de teléfono, que información académica de una universidad. En el primer caso la fuente idónea podría ser una base de datos bibliográfica o un índice, en el segundo simplemente la guía telefónica, en el tercero puede ser útil el sitio corporativo de esa universidad.

Esto que es tan evidente en nuestra vida cotidiana "real", deja de serlo, de forma absurda en Internet, donde independientemente del tipo de información que buscamos, nos solemos limitar a utilizar el Altavista, el Google o el Yahoo. Si a eso le sumamos la diferencia artificial que se ha creado entre lo real y lo virtual, la situación es aún más absurda. Cuando se habla de fuentes de información, no nos referimos exclusivamente a las que puedan encontrarse en Internet. Si absurdo es limitar la consulta a los buscadores "de turno", aún lo es más el perder el tiempo

en la red, cuando obtendríamos fácilmente lo que queremos preguntando consultando la guía telefónica.

Conocer las fuentes de información adecuadas para cada necesidad requiere de cierto tiempo, pero conociendo las más básicas y mediante un correcto aprendizaje tenemos más que suficiente para iniciarnos y resolver casi cualquier consulta.

Durante los últimos años el número de fuentes ha crecido exponencialmente. Cada página web es una fuente de información potencial. De hecho Internet es una "fuente de fuentes". No es necesario conocerlas todas. Ante cada necesidad debemos encontrar o conocer la más adecuada. Con el paso de tiempo dispondremos de una selección de la que nos son útiles (las que resuelven nuestros problemas) y que no acostumbran a ser más de 10 o 15 diferentes. Un número razonable para empezar a trabajar.

### **¿Como debo buscarlo?**

La estrategia marca, para cada fuente de información, cuestiones como, de que manera se recupera la información (navegando o interrogando a una base de datos), que palabras o conceptos se utilizan, en que idioma se formulan, etc.

Se debe adecuar la estrategia a la fuente de información que se utiliza, ya que cada una de ellas tiene sus particularidades que obligan a personalizarla. No obstante, existen normas básicas que son aplicables a la inmensa mayoría de fuentes de información y que permiten obtener resultados, sin necesidad de ser uno experto.

Dedicaremos especial atención a las estrategias para optimizar las consultas en las herramientas de búsqueda más comunes, como son los directorios y buscadores.

### **Ejecución de la estrategia**

La ejecución de la consulta no es un proceso banal. No consiste en poner una secuencia de palabras en un buscador, hacer clic sobre el botón y esperar obtener exactamente lo que buscamos. La falta de un sistema o estrategia a la hora de buscar dificulta la búsqueda.

La estrategia es la forma en la que planeamos la consulta para lograr un fin determinado

### **La interrogación de buscadores y bases de datos.**

**El lenguaje natural** es el que utilizamos las personas para comunicarnos. Los buscadores no son más que máquinas y programas que, hoy por hoy, no están

preparadas para captar las sutilezas del lenguaje humano. En consecuencia, utilizar el lenguaje natural produce resultados con un excesivo número de documentos y además, de dudosa utilidad.

### **Uso de operadores booleanos**

El álgebra de Boole proporciona tres operadores lógicos (**AND**, **OR**, **NOT**) y las reglas mediante las cuales se calcula el resultado de combinar ideas con cualquiera de ellos. Los resultados obtenidos solo pueden ser verdaderos o falsos. Mediante el Álgebra de boole se es mucho más preciso en las búsquedas y se obtiene el máximo de documentos relevantes y se elimina la infinita cantidad de irrelevantes.

Ejemplos de búsquedas booleanas:

- Edición **AND** Lineal
- Magnetoscopio **OR** VTR
- Magnetoscopio **AND** VTR **AND NOT** VHS

Las posibilidades se multiplican con el uso de paréntesis.

### **Literales**

Otra posibilidad que aportan los buscadores es el uso de literales, es decir que busque exactamente la palabra que se le indica eliminando todas aquellas que simplemente se parecen. Para indicar al buscador que busque un literal se acostumbra a poner la palabra o frase entre comillas o dobles comillas, aunque siempre está en función de la fuente de información utilizada

Ejemplo: **"Universidad Cardenal Herrera CEU"**

### **Truncamientos**

Los truncamientos son útiles para buscar, a partir de la raíz de la palabra, todos sus derivados. Eso evita tener que especificar cada una de ellas a la hora de interrogar a la base de datos.

Ejemplo: **infor\***

El sistema recuperará términos como informática, información, informatización o informativos.

### **Búsqueda por campos**

Otra de las opciones que permiten las bases de datos de los buscadores es la búsqueda en campos concretos. Los campos son diferentes en función de la base

de datos. Los más habituales en los buscadores son título, texto, fecha de publicación e idioma.

### **Formularios de búsqueda**

Las bases de datos ofrecen soluciones basadas en la presentación de formularios en los cuales los operadores están implícitos y/o desarrollados de una manera más comprensible.

Los fundamentos para realizar la consulta son los mismos que en álgebra de boole, pero el uso de formularios permite realizar la consulta de una manera menos abstracta y más comprensible al usuario medio

*Independientemente de la fuente que se consulte, la navegación y la interrogación son las técnicas que combinadas permiten recuperar información.*

### **Fuentes de información general.**

#### **Directorios o Índices.**

**Los directorios** ofrecen la información contenida en un número limitado de páginas web previamente seleccionadas por su interés y la organizan de forma jerárquica por materias, de manera que el usuario pueda descender los niveles de especificidad necesarios hasta encontrar la información adecuada a sus necesidades. En los directorios existe una selección y una clasificación de páginas Web por categorías, realizada por personas.

En el directorio, el usuario no necesita formular con términos su consulta, sino que es suficiente que seleccione las materias más afines con sus necesidades. Este sistema garantiza que todos los documentos recuperados van a tratar efectivamente del tema en el que se han clasificado, pero probablemente se hayan perdido muchos otros que no han sido seleccionados por los responsables del directorio.

Aunque los Directorios tengan incorporado un motor de búsqueda, es importante destacar que su principal valor añadido es la intervención del elemento humano en el diseño de su estructura y organización interna (la ya comentada rama jerárquica), así como en la selección y clasificación de los documentos que contiene. Esto solventa el mayor problema de Internet que es la falta de estructura, pero también limita mucho más la información recuperable. En consecuencia, si se quiere explotar al máximo sus virtudes y minimizar sus carencias, es recomendable utilizar como forma de recuperar la información la navegación y no la consulta de su base de datos.

No obstante, es recomendable utilizar la opción de búsqueda cuando se tiene dificultades en localizar la categoría en la que han quedado clasificados los

documentos. En estos casos, como paso previo, se interroga a la base de datos con una palabra que consideremos muy significativa y de esa manera veremos en qué categorías ha quedado asignada.

### **Buscadores, motores de búsqueda o robots buscadores.**

Los robots buscadores son un tipo de herramientas que utilizamos para buscar información en la red, y estos, constan de cuatro partes:

- Un **programa robot** que explora Internet, localiza documentos, sus direcciones y las incorpora a su bases de datos.
- Un **sistema automático** para analizar los **contenidos** de los documentos localizados por el robot.
- Un **sistema de interrogación** basado en la **lógica booleana** y que, mediante un lenguaje de interrogación, permite consultar al usuario sobre los documentos que más se ajustan a sus necesidades de información.
- Un programa que actúa como **pasarela** o **gateway** entre los documentos y las bases de datos.

Los criterios para escoger un buscador son los siguientes:

- La **exhaustividad**. El número de documentos que contiene su base de datos en relación al total de documentos presentes en la red (no es lo mismo buscar en una base de datos con 6 millones de documentos que en una con 60.000 documentos).
- La **especificidad** de la información que contienen los documentos de las bases de datos del buscador (no es lo mismo buscar en un buscador especializado en el sector agrícola que en otro de contenidos generales).
- La sofisticación de la **estructura de la base de datos** (no es lo mismo poder buscar documentos en todo el mundo que poder discriminar por áreas geográficas).

Las opciones que permiten utilizar el álgebra de boole no suelen aparecer en la página principal de los buscadores. El acceso ha opciones más avanzadas de búsqueda suele estar indicado con palabras como Power Search, Advanced Search, More Search Options, Búsqueda Avanzada, etc.

### **Buscadores: Funcionamiento y tipos de búsqueda (simple y avanzada).**

Los **motores de búsqueda** se encargan de organizar la gran cantidad de datos publicados en la red, de tal manera que sea posible localizar la información que usted necesita con una cierta eficacia. En pocas palabras, los buscadores son grandes bases de datos que contienen información sobre páginas web.

Son los creadores de las páginas los que deben de darse de alta en cada uno de éstos buscadores. Es por eso que no en todos se localiza la misma información y las mismas páginas. Existen empresas y programas encargadas de actualizar la información en diversos buscadores, así que no es de extrañar que haciendo la misma búsqueda desde dos páginas distintas se obtenga unos resultados absolutamente diferentes.

### BÚSQUEDA SIMPLE.

#### **Búsquedas en Lenguaje Natural.**

Generalmente los motores de búsqueda incluyen un campo dónde introducir una cadena de caracteres (texto) y un botón para iniciar la búsqueda.



Vamos a comenzar por una búsqueda en lenguaje natural, esto es, en su idioma y de la misma forma que si le hubiera preguntado a un humano.

Tecleando una palabra, o un frase, o incluso una pregunta (por ejemplo, **productoras Valencia** o **¿qué productoras hay en Valencia?**), y a continuación haciendo **click** en Buscar (o pulsando la tecla Intro/Enter). Se puede escoger el idioma en el que aparecen las páginas solo con desplegar la lista de la derecha.



AltaVista ofrecerá una o varias páginas de respuestas (hasta 10 en cada una en la configuración básica). Si lo que se busca no aparece en la primera o en la segunda de estas páginas de respuestas es recomendable añadir más términos a la búsqueda, de forma que se vuelva más específica.

En la parte inferior de la página de resultados aparece una barra que permite acceder a las demás páginas encontradas. Si se quiere ir a otra página, sólo se tiene que hacer clic sobre el número correspondiente, o bien pulsar en **Siguiente>>**.

Result Pages: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 [\[Next >>\]](#)

word count: Montevideo: 185502; restaurantes: 273853

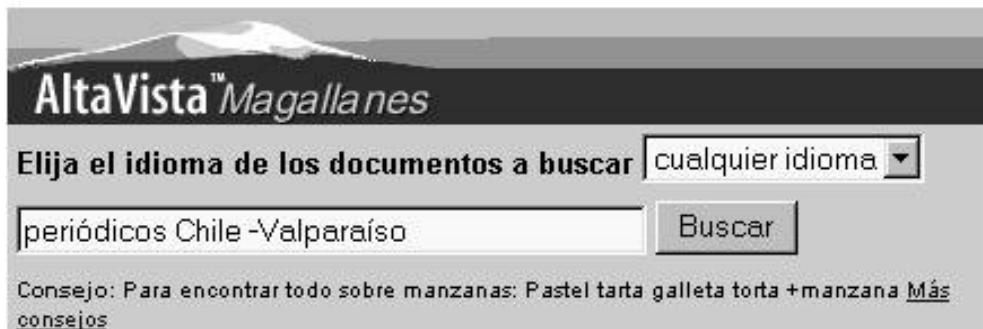
### **Términos Requeridos y Términos Excluyentes.**

Otro sistema de búsqueda es simplemente introducir términos que deben de aparecer en el resultado de la búsqueda. Tecleando las palabras una detrás de otra. Por ejemplo para encontrar productoras de videos industriales en España: **productoras vídeo industrial España.**

En las ocasiones en que se busque documentos en que algún término debe estar presente de un modo forzoso, debe identificar dicho término, incluso si es el primero, añadiéndole un signo + (por ejemplo, si busca tiendas de música en Valencia, se debe teclear **tienda música + Valencia**).

En las ocasiones en que los resultados de una búsqueda resultan excesivamente numerosos, a causa de alguna vaguedad en los términos propuestos, se puede limitar su número por medio de la exclusión de los documentos que contengan algún término en el que no se está interesado (por ejemplo, si se busca periódicos o editoriales en Francia, pero no se tiene interés en ninguno de Toulouse, se debe teclear **periódicos editoriales Francia -Toulouse**).

Hay que tener cuidado de separar los signos con espacios (no es lo mismo hospital + clínica que hospital + clínica ya que el sistema entiende hospital + clínica como hospital clínica)



### **Frases literales.**

Si se está seguro de que una determinada frase forma parte del documento o documentos en que se está interesado, se debe teclear precedida y seguida por el símbolo dobles comillas (por ejemplo, si busca documentos que hagan referencia a una determinada obra del poeta Machado, teclee **"Mi infancia son recuerdos de un patio de Sevilla"**).



## Mayúsculas/minúsculas.

En general es recomendable usar solamente letras minúsculas, incluso para nombres. De hacerse así AltaVista ofrece resultados en los que no diferencian entre mayúsculas y minúsculas.

Si se emplean mayúsculas los resultados no incluirán los que estén realizados en minúsculas (ejemplo: el término **+ gabriela sabatini** localiza todos los documentos que hablen de gabriela sabatini, de Gabriela Sabatini y de GABRIELA SABATINI; por el contrario, el término **+ Gabriela Sabatini** sólo encontraría los que hablen de Gabriela Sabatini).

## Comodines.

El empleo de comodines incrementa la potencia y la facilidad de búsqueda. Es particularmente cómodo en el caso de singular/plural (ejemplo: **+museo +museos** equivale a **+museo\***), y también en el de términos que poseen una parte común y que en principio no son fáciles de discriminar (ejemplo: **auto\*** permite localizar autobús, automoción, autostrada, autobahn, autoroute, automatismo, Autobianchi, automotor y Autonomías, entre muchos otros)

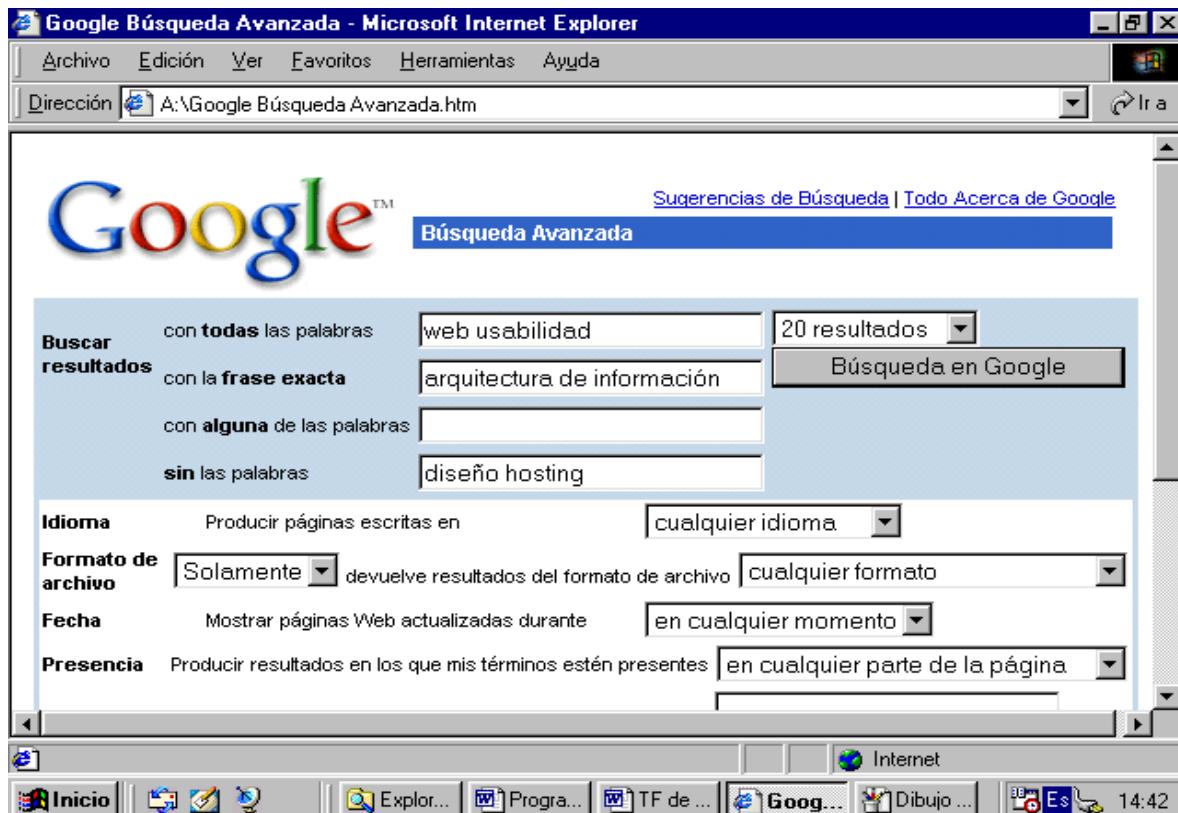
## BÚSQUEDA AVANZADA.

### Operadores Booleanos.

**Operador OR:** Localiza documentos que contengan al menos uno de los términos especificados (ejemplo: **Garrincha OR Pelé OR Vavá** localiza todos los documentos que citen al menos a uno de estos tres futbolistas brasileños; algunos de estos documentos puede que citen a dos de ellos, y algunos otros citarán a los tres, pero sólo puede contarse con que al menos citarán a uno de los tres).

**Operador NOT:** Excluye los documentos que contengan el término Especificado (ejemplo: **Jorge Luis Borges AND NOT Pablo Neruda** localizará los documentos en que simultáneamente se cite a Jorge Luis Borges y no se cite a Pablo Neruda; el operador **NOT** no puede usarse en modo aislado, sino siempre en conjunción con alguno de los otros operadores, como en este ejemplo).

**Operador NEAR:** Localiza los documentos en que los dos términos especificados aparezcan, al menos en algún caso, separados por no más de diez palabras de texto (ejemplo: **Garibaldi NEAR Cavour** encontrará documentos en que se haga referencia a ambos y sea probable que los dos coincidan en un mismo párrafo, de forma que el contexto general tienda a referirse a ambos y no a cada uno por separado).



### Entornos (rangos) de fechas.

El objeto de esta función es permitir la de limitación de la Búsqueda Avanzada de forma que sólo se extienda a los documentos que hayan sido publicados o modificados por última vez en un entorno específico de tiempo.

Las fechas deben introducirse en el formato **dd/mmm/aa**, donde **dd** es el día dentro del mes, **mmm** son las tres primeras letras del nombre del mes y **aa** son los dos últimos dígitos del año (ejemplo: **23/ene/98**, o **23/jan/98**). Si se omite el año AltaVista-Magallanes asume que la fecha corresponde al año en curso. Si también se omite el mes se interpreta que la fecha corresponde al mes en curso.

### Rango de resultados ('ranking').

Los términos que se introducen en la casilla **ranking** sirven para complementar los resultados de la Búsqueda Avanzada. Si se repite algún término de los empleados en la búsqueda, los resultados que se ofrecen en primer lugar son los que contienen estos términos (ejemplo: **Olympic or Titanic or Britannic**, de **1/ene/97** a **23/mar/98**, permite localizar todos los documentos que hagan referencia a cuando menos uno de estos tres barcos y que hayan sido publicados o modificados por última vez entre el 1 de enero de 1997 y el 23 de marzo de 1998; si se incluye Britannic en ranking los que hagan referencia a éste serán los que parezcan primero; si en vez de incluir **Britannic** en **ranking** se incluye **White Star Line or Cunard Line** los documentos localizados serán los que hagan referencia a cualquiera de los tres barcos y a una o a las dos compañías navieras, apareciendo primero las que hagan referencia a la White Star Line).

### **Términos Compuestos.**

Es posible componer/agrupar términos en la Búsqueda Avanzada. Se hace mediante el empleo de paréntesis. Así, para buscar documentos que hagan referencia a Larry Ellison, a Scott McNealy y a Bill Gates, un buen ejemplo de Búsqueda Avanzada sería el siguiente: **((Larry or Lawrence) NEAR Ellison) and (Scott near McNealy) AND ((Bill OR William) near Gates)**.

### **Entonces, ¿búsqueda avanzada o búsqueda simple?**

La Búsqueda Avanzada es más apropiada para los casos específicos que para la búsqueda general. En la mayoría de los casos el usuario encuentra lo que busca con más facilidad, más rápidamente y con mejores resultados a través de la Búsqueda Simple, donde AltaVista controla el ranking o aparente grado de adecuación entre lo que se busca y la información disponible.

Sin embargo, si el usuario intenta localizar documentos generados en un determinado entorno de fechas, o si le es más cómodo encontrarlos a partir de operaciones **booleanas** de alguna complejidad, la funcionalidad de la Búsqueda Avanzada no sólo le resultará de gran utilidad, sino que difícilmente encontrará en Internet nada de potencia equivalente.

En Búsqueda Avanzada es el usuario quien controla el **ranking**, lo que debe tenerse bien presente, ya que si el **campo** correspondiente a **ranking** se deja en blanco los resultados aparecerán en cualquier orden, y no en el que al usuario le pudiera convenir.

### **Directorios versus Robots buscadores.**

Ni motores, ni directorios, son la solución ideal para asegurar una correcta y completa recuperación de información en la web. Por el momento, la opción más satisfactoria es la combinación de ambas herramientas.

- Para **búsquedas muy concretas** y definidas, es recomendable el uso de un **motor de búsqueda**, ya que hará una búsqueda exhaustiva en todo el web.
- Si por el contrario, lo que se quiere localizar es información acerca de un **tema en general**, conviene comenzar la búsqueda por un **directorio**, ya que proporcionará documentos relevantes sobre dicho asunto y en esas páginas se encontrarán enlaces a otras con información relacionada.

### **Multibuscadores.**

Los multibuscadores son herramientas capaces de realizar **búsquedas simultáneas en más de un buscador**. Aunque útiles, hay que tener en cuenta, que las posibilidades de utilizar operadores están limitadas. El peligro, una vez más, en encontrar un número excesivo de documentos no relevantes. Copernic tiene la particularidad de permitirnos lanzar la consulta desde nuestro ordenador, mediante el programa cliente que puede bajarse des de sus páginas.

### **Directorios temáticos.**

Los buscadores y directorios generalistas del tipo Google, Altavista, Yahoo, se muestran a menudo inútiles, cuando de lo que se trata es de encontrar información especializada. Es en estas circunstancias, cuando resulta imprescindible el uso de directorios temáticos.

Los directorios temáticos son páginas de referencia **especializadas en áreas temáticas** concretas, y son puntos de salida a partir de los cuales, localizar otros recursos aún más específicos o, en algunos casos, lanzar las consultas directamente.

### **Metabuscadores.**

Los metabuscadores son **buscadores de buscadores**. Es el recurso principal a partir del cual podemos localizar la fuente que más se ajusta, temáticamente, a nuestra necesidad de información y, a partir de ella, iniciar la que será la búsqueda real de los documentos que necesitemos.

Los metabuscadores son ideales para **encontrar el directorio temático** de referencia en la materia de la consulta que queremos realizar. Podríamos decir, que los metabuscadores son los directorios temáticos, cuya área de conocimiento son, justamente, las herramientas de búsqueda de información en Internet.

Existe una gran variedad de buscadores e índices en Internet para localizar aquella información que se busca, que, con una correcta utilización, pueden ser de gran eficacia. Aún así existen problemas estructurales (sobre todo en los buscadores) que hacen muy penosa y larga la búsqueda de contenidos cuando el usuario no tiene una idea concreta de la información que necesita. Solo la

introducción de conceptos claves muy concretos permite localizar con rapidez contenidos adecuados.

Una solución a este problema estructural, que no comentamos en este trabajo, son las bases de datos especializadas que permiten encontrar justamente aquellas palabras claves que son necesarias para una búsqueda eficaz en directorios y buscadores. Un ejemplo bueno es [www.imdb.com](http://www.imdb.com), base de datos muy completa sobre cine y televisión.

### **¿He encontrado lo que buscaba? ¿La información que aporta es útil?**

La valoración de resultados es totalmente subjetiva y es cada usuario el que debe decidir si la información obtenida es la que va a ayudarle a resolver el problema planteado.

Aun así, existen criterios, orientados normalmente a valorar la fuente de información y que son muy útiles para hacer una estimación de los resultados de una manera más objetiva.

- **La Fiabilidad** de la institución o persona responsable de su publicación.
- **La Objetividad** de la información. El fin con el que ha sido creada la información. No es lo mismo información claramente técnica que información publicada con fines comerciales.
- **El Contenido real de información**: Un fenómeno habitual en Internet es la presencia de páginas que tan solo aportan enlaces a otras páginas y en las que el contenido real de información es muy limitado.
- **La Originalidad de la información**. La información recuperada pertenece en exclusiva a la entidad que hemos consultado o puede ser obtenida en otras fuentes.
- **La Integridad de los contenidos**. Son parciales o están completos. Un ejemplo típico de contenidos parciales son la mayoría de bases de datos bibliográficas en las que los resultados no son documentos completos, sino la referencia a artículos publicados en la prensa.
- **La Actualización y Vigencia de la información**.

En conclusión, Internet es una fuente de información casi inagotable, pero solo el conocimiento de las herramientas que ofrece y su correcta utilización permite acceder a la información adecuada.

**Universidad Cardenal Herrera CEU, Valencia - España**  
**Julián Gómez Ayora & Daniel Fernández Jutz**  
**Búsqueda y recuperación de información en Internet**

<http://webs.ono.com/usr019/D.FERNANDEZ.JUTZ/introduccion.htm>

Este trabajo tiene como principal objetivo el proporcionar los conocimientos y los recursos de Internet que ha revolucionado la capacidad de acceso y difusión de la información. Pero su inmenso potencial también ha convertido la búsqueda y recuperación de información, en un problema de difícil solución. Para poder explotar los recursos presentes en Internet es necesario conocer qué fuentes de información están disponibles y qué técnicas son más necesarias para explotarlas adecuadamente.

La forma más eficaz de encontrar información en Internet es utilizando los llamados "motores" de búsqueda[1] o buscadores[2]. Sin lugar a dudas el más popular y visitado es *Yahoo!*, seguido de cerca por *Altavista*, *Excite*, *InfoSeek* y *Web Crawler*. Existen buscadores en casi todos los idiomas del mundo y algunos de ellos tienen ciertas "especialidades". Otros se limitan a páginas en un idioma determinado o referentes a un cierto país o región.

Para obtener la información, las empresas propietarias de los diferentes buscadores invierten una gran cantidad de tiempo y potencia de sus sistemas de computación para buscar, levantar información y catalogarla por criterios lógicos. La otra forma de obtener datos es mediante la información que envían los propios responsables de los "sites".

Además de los "grandes" buscadores clásicos y "universales" de Internet, hay toda una gama de buscadores medianos, que apuestan por aspectos más particulares: es el caso de los buscadores que se especializan en un idioma, o en un país, o en una temática, etc.

Los criterios de montaje varían también de un buscador a otro. Algunos funcionan a partir de robots que indexan automáticamente los sitios que se les proponen. Otros en cambio incluyen una operación personalizada en la indexación. Esto también puede tener importancia a la hora de buscar, pues en los primeros la información será mayor (se encontrarán indexadas casi todas las páginas de cada sitio) pero también más desordenada, en cambio en los segundos la información será menor (solo se encontrarán las páginas de entrada de cada sitio) aunque, en principio, de modo más ordenado.

Internet es utilizado en la mayoría de los casos en busca de entretenimiento, como medio de comunicación directo entre personas o para la búsqueda y recuperación de información. Es este último caso el que trataremos en este trabajo.

[1] Un motor de búsqueda es un programa que busca a través de un conjunto de datos. En el contexto del Web, la palabra 'motor de búsqueda' se usa para referirse a programas que buscan en una base de datos de documentos HTML indexados por un robot.

[2] Los buscadores son máquinas que se encargan de registrar todas las direcciones URL que circulan por la web cada cierto tiempo. El usuario puede pedirle cualquier palabra clave y ellos le devolverán las direcciones de todas las páginas que la contienen la formación necesaria para mejorar la actividad de búsqueda de información mediante el uso de Internet. A lo largo de los siguientes puntos iremos explicando los pasos que se han de seguir cuando tengamos una necesidad de información, qué fuentes de información son más adecuadas para cada caso y qué técnicas nos permiten optimizar nuestras consultas en cada caso, y todo ello mediante el empleo de los llamados buscadores de Internet.

**Con respecto a Internet, hay que tener en cuenta lo siguiente:**

- Internet no es la única fuente.
- Buscar información en Internet no es una operación de una sola etapa.
- Internet no es un producto comercial, sino la suma de miles de voluntades individuales, con todas las ventajas e inconvenientes que esto supone.
- Internet aún no es rápida.
- Internet no es una biblioteca universal perfectamente organizada.
- En Internet no existe "a priori" ningún control de calidad de los contenidos.
- Las fuentes de información en Internet no son estáticas o estables.
- Internet no es la respuesta.

## LOS MEJORES BUSCADORES

1. AltaVista: <http://altavista.com/>
3. Excite: <http://excite.com/>
5. Go (antes, Infoseek):  
<http://go.com/>
7. Google: <http://google.com/>
9. Lycos: <http://lycos.com/>
11. Snap: <http://www.snap.com/>
2. Direct Hit: <http://www.teoma.com/>
4. Fast: <http://alltheweb.com/>
6. GoTo: <http://www.goto.com/>
8. HotBot: <http://www.hotbot.com/>
10. Northern Light:  
<http://www.northernlight.com/>

## ARGENTINA

- . Clarín: <http://www.buscador.clarin.com.ar/>
3. Radar: <http://www.radar.com.ar/>
2. Grippo: <http://www.grippo.com/>
4. Yahoo! Argentina: <http://ar.yahoo.com/>

## BRASIL

5. Achei: <http://www.achei.com.br/>
7. Cadê?: <http://www.cade.com.br/>
9. UOL: <http://www.uol.com.br/>

## **BOLIVIA**

11. Bolivian: <http://www.bolivian.com/>

## **CHILE**

12. AgendaChile: <http://agendachile.com/>
14. Buscar.cl: <http://www.buscar.cl/>
16. Chileinternet: <http://www.chileinternet.cl/>
18. Cronos: <http://www.cronos.cl/>
20. El Faro: <http://www.faro.cl/>
22. La Brújula: <http://www.brujula.cl/>
24. TodoCL: <http://www.todocl.cl/>
26. P. Amarillas: <http://www.paginasamarillas.cl/>
13. Antena: <http://www.antena.cl/>
15. Chile en Internet: <http://www.guia-chile.cl/>
17. Chile Online <http://www.chile-online.com/>
19. Domo: <http://www.domo.cl/>
21. Huifa: <http://www.huifa.cl/>
23. Mercantil.com: <http://www.mercantil.com/>
25. OpenChile: <http://www.openchile.cl/>

## **COLOMBIA**

27. Conexcol: <http://www.conexcol.com>
28. eHOLA.com: <http://www.ehola.com>

## **COSTA RICA**

29. Orientation Costa Rica: <http://www.orientation.co.cr>

## **ECUADOR**

30. ¡Mande?: <http://www.mande.com.ec>
31. Webinator: <http://www.ecuasearch.net.ec>

## **ESPAÑA**

33. Elcano: <http://www.elcano.com>
35. Ozú: <http://www.ozu.com>
37. Yahoo! España: <http://es.yahoo.com>
34. Hispavista: <http://www.hispavista.com>
36. Terra: <http://www.terra.es>

## **LATINOAMÉRICA**

38. Astrolabio: <http://www.astrolabio.net>
40. El Buscador: <http://www.elbuscador.com>
42. LatinGuia: <http://www.latinguia.com>
44. StarMedia: <http://www.starmedia.com>
39. BUSCApique: <http://www.buscapique.com>
41. El Sitio: <http://www.elsitio.com>
43. Metabusca: <http://www.metabusca.com>
45. Yupi: <http://www.yupi.com>

46. Terra: <http://www.terra.es>

48. Yahoo! en Español:  
<http://espanol.yahoo.com>

47. UOL: <http://www.uole.com>

49. TeRespondo.com: <http://www.terespondo.com>

## **MÉXICO**

50. Adnet: <http://www.adnet.com.mx>

52. MéxicoGlobal: <http://www.mexicoglobal.com>

54. México Web: <http://www.mexicoweb.com.mx>

51. Mexmaster: <http://www.mexmaster.com>

53. Yahoo! México: <http://mx.yahoo.com>

## **PARAGUAY**

55. Yagua: <http://www.yagua.com>

**PERÚ** 56. Adónde?: <http://www.adonde.com>

57. Nazcanet: <http://www.nazcanet.com>

## **URUGUAY**

58. Uruguay Total: <http://www.uruguay.org.uy>

## **VENEZUELA**

59. Auyantepui: <http://www.ayuantepui.com>

60. Croquer: <http://www.croquer.com>

61. Venezuelasite: <http://www.venezuelasite.com>

## **BUSCADORES EUROPEOS**

### **ALEMANIA**

1. DINO-Online: <http://www.dino-online.de>

3. WEB.DE: <http://web.de>

2. Fireball: <http://www.fireball.de>

4. T-Online: <http://www.t-online.de>

### **AUSTRIA**

5. T-Online: <http://www.t-online.at>

### **BÉLGICA**

6. Belcast: <http://www.belcast.be>

7. WebWatch: <http://www.webwatch.be>

### **DINAMARCA**

8. DEnet: <http://info.denet.dk>

9. Jubii: <http://www.jubii.dk>

## **EUROPA**

10. EuroSeek: <http://www.euroseek.com>  
12. Search Europe:  
<http://www.searcheurope.com>
11. WebTop.com: <http://www.webtop.com>  
11. Orientation Central & Eastern Europe:  
<http://eeu.orientation.com>

## **FRANCIA**

14. Carrefour.net: <http://carrefour.net>  
16. Écila: <http://www.ecila.fr>  
18. Lokace: <http://www.lokace.com>  
20. Voila: <http://www.voila.fr>
15. Club-Internet: <http://www.club-internet.fr>  
17. Eurêka, c'est trouvé!: <http://www.eureka-fr.com>  
19. Nomade.fr: <http://www.nomade.fr>  
21. Wanadoo: <http://www.wanadoo.fr>

## **HOLANDA (PAÍSES BAJOS)**

22. NL.Menu: <http://www.nl-menu.nl>  
24. Track: <http://www.track.nl>
23. Search.NL: <http://www.SEARCH.nl>  
25. Zoek.nl: <http://www.zoek.nl>

## **HUNGRÍA**

26. Heuréka: <http://www.heureka.hu>
27. HuDir: <http://hudir.hungary.com>

## **INGLATERRA (REINO UNIDO)**

28. UK Plus: <http://www.ukplus.com>  
30. Searchengine.com:  
<http://uk.searchengine.com>
29. UK Directory: <http://www.ukdirectory.com>

## **ISLANDIA**

31. The Web Collection: <http://www.hugmot.is/ssafn/english>

## **ITALIA**

32. Arianna: <http://arianna.iol.it>  
34. Virgilio: <http://www.virgilio.it>
33. IlTrovatore: <http://www.iltrovatore.it>

## **MALTA**

35. Search Malta: <http://www.searchmalta.com>

## **NORUEGA**

36. Scandinavia Online: <http://www.sol.no>

## RUMANIA

37. Cauta.ro: <http://www.cauta.ro>

## RUSIA

38. Russia on the Net: <http://www.ru>

## SUIZA

39. Search.ch: <http://www.search.ch>

## DIRECTORIOS

- |  |  |
|--|--|
| 1. Yahoo!: <a href="http://yahoo.com">http://yahoo.com</a>                       | 2. About.com: <a href="http://about.com">http://about.com</a>  |
| 3. LookSmart: <a href="http://www.looksmart.com">http://www.looksmart.com</a>    | 4. Snap: <a href="http://www.snap.com">http://www.snap.com</a> |
| 5. Open Directory Project: <a href="http://www.dmoz.org">http://www.dmoz.org</a> |  |

## METABUSCADORES

- |   |   |
|---|---|
| 1. MetaCrawler: <a href="http://www.metacrawler.com/">http://www.metacrawler.com/</a> | 2. Mamma: <a href="http://www.mamma.com/">http://www.mamma.com/</a>             |
| 3. Savvy.com: <a href="http://www.savvy.com/">http://www.savvy.com/</a>               | 4. Dogpile: <a href="http://www.dogpile.com/">http://www.dogpile.com/</a>       |
| 5. All4one: <a href="http://www.all4one.com/">http://www.all4one.com/</a>             | 6. Highway61: <a href="http://www.highway61.com/">http://www.highway61.com/</a> |
| 7. Metabusca: <a href="http://www.metabusca.com/">http://www.metabusca.com/</a>       | 8. Antena: <a href="http://www.antena.cl/">http://www.antena.cl/</a>            |

## El modelo Big6 para la solución de problemas de información

Por: Mike Eisenberg y Bob Berkowitz

### El Modelo Big6 para la solución de Problemas de Información

El Modelo Big6 (Los Seis Grandes 1) desarrollado por Mike Eisenberg y Bob Berkowitz se puede definir como un proceso sistemático de solución de problemas de información apoyado en el pensamiento crítico. También podría definirse como las seis áreas de habilidad necesarias para la solución efectiva y eficiente de problemas de información (puntos específicos y estratégicos que ayudan a satisfacer las necesidades de información) o como un currículo completo de habilidades para el uso de la biblioteca y el manejo de la información. Las habilidades tradicionales para usar la biblioteca se enfocan en el conocimiento y la

comprensión de fuentes específicas (habilidades cognitivas de orden inferior), en contraposición con la habilidad de utilizar el pensamiento crítico y la capacidad de manipular la información para lograr soluciones significativas.

Los estudiantes pueden utilizar esta herramienta cuando se vean en una situación académica o personal que requiera información precisa para resolver un problema, tomar una decisión o realizar un trabajo. Mediante el empleo de un enfoque de niveles múltiples, los estudiantes pueden desarrollar competencia tanto en la solución de problemas de información como en la toma de decisiones, lo que más adelante se convertirá para ellos en una habilidad permanente.

Big6™ consta de los siguientes pasos:

1	Definición de la Tarea a Realizar	Enfocar	¿Cuál es el problema?
2	Estrategias para Buscar Información	Planear la Búsqueda	¿Cómo debo buscar?
3	Localización y Acceso	Clasificar	¿Qué obtuve?
4	Uso de la Información	Seleccionar	¿Qué es lo importante?
5	Síntesis	Sintetizar + Producir	¿Cómo encajan juntos? ¿A quién va dirigido?
6	Evaluación	Evaluuar, Reflexionar	Y entonces, ¿Qué aprendí?

Una forma eficiente de implementar las habilidades del modelo Big6 es buscar oportunidades entre las lecciones o temas académicos de clase que estén relacionados con estas habilidades para aplicarlas. Ver artículo sobre la aplicación práctica del Big6. Dentro de un ambiente en el que se busque desarrollar la competencia en el manejo de la información, los estudiantes se involucran en actividades de aprendizaje dinámicas, autodirigidas, y los profesores facilitan el proceso identificando fuentes y recursos por fuera de la clase. Cuando el estudiante entra a la biblioteca, ya sospecha que existe un problema de información; pero en ese punto, casi nunca ha identificado lo que le falta. Es en ese preciso momento cuando las ventajas del modelo Big6 se evidencian.

## **1. DEFINICIÓN DE LA TAREA: ¿Cuál es el problema?**

En esta etapa del proceso se trata de determinar exactamente cuál es el problema de información que existe y de definir las necesidades específicas relacionadas con este. Por ejemplo, en el caso de una tarea escolar los estudiantes deben saber cuales son las preguntas que se deben responder y la clase de información necesaria para contestar dichas preguntas.

¿Cuál es el problema por resolver?

¿Qué información se necesita para solucionar el problema?

¿Qué es lo que se requiere para realizar la tarea?

¿En qué orden se debe proceder y cuál es el tiempo máximo para completar los pasos a seguir?

**Ejemplos de Definición de la Tarea:**

- Crear un diagrama del proceso de solución de un problema específico.
- Comparar el enfoque usual para resolver problemas matemáticos con el método que propone Big6 para tal fin.
- Enumerar los pasos para finalizar una tarea; luego, compararlos con los del método Big6.
- Mencionar algunos casos en que se usa información para resolver problemas.
- Especificar si algunas de las actividades a desarrollar en el laboratorio de ciencias requieren información específica.
- Identificar las necesidades de información del problema. Para cada una de las actividades relacionadas con información en el mismo ejercicio de laboratorio, anote si requieren localización y acceso, uso de información o síntesis de esta.
- Definir si la tarea hace necesario tanto rotular como mirar un mapa o croquis.
- Definir los pasos a seguir en una clase de gimnasia y anotar si se necesita alguna información para cada uno de ellos.

**2. ESTRATEGIAS DE BÚSQUEDA DE INFORMACIÓN: ¿Cómo debo buscar?**

Cuando el problema esté claramente definido, se analiza la gama de posibles fuentes de información. Las estrategias para buscar información implican tomar decisiones y escoger las fuentes de información más convenientes para el trabajo en cuestión.

¿Cuales son las posibles fuentes de información?

¿Cuales son las mejores de estas posibilidades?

¿Cuales son los métodos alternos para obtener información?

**Ejemplos de Búsqueda de Información:**

- Con una lluvia de ideas proponer las formas de encontrar información sobre Elvis Presley.
- Decidir cuales pueden ser las fuentes de referencia para encontrar información sobre deportistas famosos.
- Enumerar las fuentes donde se puede encontrar información sobre crítica literaria.
- Hacer un inventario de todos los recursos de computación que hay en el colegio.
- Evaluar los posibles recursos para establecer prioridades.

- Decidir si se le pregunta a un experto o si se busca en un texto de referencia o en alguna otra fuente.
- Decidir si está permitido consultar una enciclopedia para hacer un trabajo.

### **3. LOCALIZACIÓN Y ACCESO: ¿Qué obtuve?**

En este punto empieza la investigación propiamente dicha. Después de que los estudiantes hayan definido la estrategia más apropiada a seguir, ésta se debe poner en práctica. Esta es la parte física y la que recibe mayor atención en el programa curricular de una biblioteca tradicional. Incluye: uso de herramientas de acceso, ordenamiento de materiales, partes de un libro y estrategias para buscar en un catálogo en línea, en bases de datos o en Internet. Con mucha frecuencia las instrucciones que se dan en una biblioteca están enfocadas hacia el uso de habilidades específicas asociadas con herramientas especiales de búsqueda, como catálogos o motores de búsqueda, en lugar de centrarse en habilidades que se puedan transferir a otras situaciones o a otros problemas. Con el enfoque de Big6, se sigue el proceso lógico de decidir primero que es lo que se quiere buscar y donde; para definir posteriormente los materiales necesarios:

- ¿Dónde están estas fuentes?
- ¿Dónde está la información al interior de cada fuente?

#### **Ejemplos de Localización y Acceso a la Información:**

- Conseguir un artículo de revista en la biblioteca, buscar la revista correspondiente y llegar a la página correcta del artículo en cuestión.
- Ir a la biblioteca pública y consultar un libro sobre un presidente Colombiano.
- Localizar las fuentes (intelectual y físicamente):
  1. Encontrar un libro específico en las estanterías.
  2. Dibujar y rotular un plano de la biblioteca.
- Encontrar información al interior de las fuentes:
  1. Buscar un artículo sobre una serie de televisión conocida2.
  2. Encontrar un artículo sobre música rock usando un índice periódico en un CD-ROM.

### **4. USO DE LA INFORMACIÓN: ¿Qué es lo importante?**

Una vez que los estudiantes sean capaces de localizar y acceder una fuente, deberán poder leer, visualizar, escuchar e interactuar con la información disponible y decidir que aspectos de ésta son útiles para su situación específica. Deberán extraer la información necesaria tomando notas, sacando copias, anotando citas, etc.

¿Qué información ofrece la fuente?

¿Qué información específica vale la pena utilizar para el trabajo que se va a realizar?

### **Ejemplos de Uso de Información:**

- Ver un video sobre terremotos e identificar los puntos más importantes.
- Estudiar el glosario en la parte de atrás de un libro en busca de un término específico y si aparece, copiar la definición.
- Usar la información en una fuente (leerla, verla, escucharla, etc.):
  1. Examinar rápidamente un libro para saber si puede ser útil
  2. Escuchar un casete de audio de un libro famoso, por ejemplo "Historia de Dos Ciudades".
- Extraer información de una fuente:
  1. Tomar notas sobre información bibliográfica para usarlas posteriormente.
  2. Tomar notas del artículo de una revista.

### **5. SÍNTESIS: ¿Cómo encaja la información? ¿A quién va dirigida?**

La síntesis es la reestructuración o el reempaque de la información en formatos nuevos o diferentes para poder cumplir con los requisitos del trabajo. La síntesis puede ser tan simple como trasmitir un hecho específico o lo bastante compleja, como para contener varias fuentes, varios formatos de presentación o diferentes medios de información y la comunicación efectiva de ideas abstractas.

¿Cómo se une la información que proviene de distintas fuentes?

¿Cómo se presenta mejor la información?

### **Ejemplos de Síntesis:**

- Hacer un resumen (usando información de fuentes múltiples) para un informe.
- Preparar una producción de video con algunas de las facilidades que haya en el colegio.
- Organizar la información obtenida de fuentes múltiples:
  1. Crear una base de datos sobre las principales ciudades de América Latina.
  2. Organizar tarjetas con notas (de fuentes múltiples) en orden lógico.
- Presentar la información:
  1. Hacer una copia de una base de datos.
  2. Dibujar y rotular un mapa de África.

### **6. EVALUACIÓN: Y entonces... ¿qué aprendí?**

La evaluación es la que determina qué tan efectiva y eficientemente se llevó a cabo el proceso de solución del problema de información. Los siguientes son los puntos más importantes que se deben evaluar:

- ¿El problema de información quedó resuelto?
- ¿Se obtuvo la información que se necesitaba?
- ¿Se tomó la decisión?
- ¿Se resolvió la situación?
- ¿El producto obtenido cumple con los requisitos originalmente establecidos?
- ¿Qué criterios se puede utilizar para hacer evaluaciones?

Otros aspectos que se deben tener en cuenta para evaluar la efectividad del proceso de solución del problema, incluyen la cantidad de tiempo empleado en actividades útiles y si hubo algún error de cálculo en la cantidad de tiempo requerido para finalizar las tareas. Esta auto evaluación por parte del estudiante mejorará su habilidad futura para resolver con mayor facilidad otros problemas de información.

#### **Ejemplos de Auto-Evaluación:**

- Encontrar las razones por las cuales el estudiante no obtuvo A en su proyecto.
- Decidir si un trabajo ha quedado finalizado o no.
- Juzgar la efectividad del producto final:
  1. Definir los criterios para evaluar los avisos que prohíben fumar.
  2. Establecer si ha quedado satisfecha la necesidad de información de acuerdo a lo planteado inicialmente.
- Evaluar la efectividad del proceso de la solución del problema de información:
  1. Definir el grado de efectividad de la técnica para tomar notas.
  2. Explicar que cambiaría en el procedimiento la próxima vez.

Aunque el proceso de Big6 siempre se presenta como si se debiera realizar paso a paso (Desde "Identificación de la Tarea" hasta "Evaluación"), generalmente las personas no lo siguen así. La solución exitosa del problema de información requiere completar adecuadamente en algún momento, cada una de las etapas del modelo Big6 pero las personas saltan de una etapa a otra desordenadamente, retroceden o se desvía. El énfasis se debe hacer en desarrollar competencia y pericia en cada una de las áreas del modelo Big6, sin que esto signifique una camisa de fuerza.

# **Algunas herramientas de búsqueda en Internet. Su evolución y estado actual.**

***Dra. Gloria Hernández Muguerzia***

***Dra. Mirta Valdés Morris***

## **Resumen**

Lo vertiginoso del desarrollo de las herramientas de búsqueda en la Red nos obliga a mantenernos al día en los cambios que se van produciendo en este vasto campo. Este trabajo pretende aportar información actualizada sobre lo que está ocurriendo en el ambiente Internet en este sentido, y la conveniencia de estar al tanto de la evolución y consolidación de algunos de los más populares buscadores, entre ellos Yahoo, Altavista, Excite, Lycos y otros.

Se presentan algunos indicadores de efectividad en el uso, representatividad y cubrimiento del Web, que nos aportan elementos acerca de cuán eficientes pueden ser unos con respecto a los otros, y de hacer la mejor elección entre ellos, para obtener resultados de búsquedas más pertinentes, ello favorece el trabajo del profesional de la información y a todo aquel que necesita encontrar lo que busca, de la mejor manera posible y con un considerable ahorro en la utilización de su tiempo en línea. Se trata también sobre la tendencia de agregar valor a estos sitios con el surgimiento de un reciente concepto: los llamados "portales" y de la influencia que estos tienen en servicios cada vez más personalizados al usuario final, entre otros beneficios.

## **Introducción**

Lo vertiginoso del desarrollo de las herramientas de búsqueda en la Red nos obliga a mantenernos al día en los cambios que se van produciendo en este vasto campo. Este trabajo pretende aportar información actualizada sobre lo que está ocurriendo en el ambiente Internet en este sentido, y la conveniencia de estar al tanto de la evolución y consolidación de algunos de los más populares buscadores, entre ellos Yahoo, Altavista, Excite, Lycos y otros.

Se presentan algunos indicadores de efectividad en el uso, representatividad y cubrimiento del Web, que nos aportan elementos acerca de cuán eficientes pueden ser unos con respecto a los otros, y de hacer la mejor elección entre ellos, para obtener resultados de búsquedas más pertinentes, ello favorece el trabajo del profesional de la información y a todo aquel que necesita encontrar lo que busca, de la mejor manera posible y con un considerable ahorro en la utilización de su tiempo en línea.

Se trata también sobre la tendencia de agregar valor a estos sitios con el surgimiento de un reciente concepto: los llamados "portales" y de la influencia que

estos tienen en servicios cada vez más personalizados al usuario final, entre otros beneficios.

De las autopistas de la información actuales, sigue siendo Internet la más conocida y transitada, ello con muy diversos fines: aprendizaje, entretenimiento, actualización de noticias, adquisición de cultura general, entre otros, pero según encuesta de la publicación Business Week, la mayor parte de los usuarios de la red consumen el 50% de su tiempo en línea haciendo búsquedas.

Son múltiples las herramientas de las cuales los usuarios pueden servirse para lograr este objetivo, no obstante, centraremos este trabajo en los buscadores y diferenciaremos en estos a los Directarios Web, los Motores de Búsqueda y a los Metabuscadores, habida cuenta de que no pretendemos sentar pautas conceptuales al respecto, pero sí revisar lo que especialistas muy calificados tienen en cuenta sobre este tema y las habilidades y experiencias adquiridas en el trabajo diario durante algún tiempo.

Como su nombre lo indica, los buscadores nos facilitan "buscar" y encontrar información diseminada por la red, con la particularidad de que cada uno dará sus resultados según la información que abarca su propia base de datos o su listado de direcciones y de acuerdo con los criterios de indexación que en estos se mantenga, y no como erróneamente se piensa, al suponerse que ha sido explorado todo el Web, y por ende todas las páginas en él registradas.

Como consecuencia lógica, todos los buscadores ofrecen la facilidad de búsqueda, pero hay algunos elementos que han de tenerse en cuenta, para poder elegir cuál de ellos nos devolverá los mejores resultados y en el menor tiempo posible, partiendo de un buen criterio o estrategia de búsqueda en nuestra solicitud, cuestión que no puede ser eludida.

Para empezar, es necesario conocer algunas de las diferencias que existen entre un directorio Web y un motor de búsqueda:

### **Directorio Web:**

- es una lista predefinida de sitios Web
  - es compilado por personas
  - está categorizado por tópicos o materias
  - devuelve enlaces relacionados y útiles

Los directorios Web, como hemos dicho antes, son compilados por personas (equipo de expertos), quienes deciden, después de analizar las Uniform Resources Locator (URL) que reciben y sus contenidos, qué sitios serán incluidos en su lista predefinida. Precisamente por ser el recurso humano más costoso y más lento que las computadoras, los directorios sólo pueden cubrir un pequeño

por ciento de lo que está disponible en la red, por ejemplo, el mayor de ellos, Yahoo, sólo cubre alrededor de un 1,2 millones de sitios Web, otros también muy conocidos son Lycos, Magellan, UK Directory, Olé, Ozú, Informed Business Services, Business Researcher's Interests, Look Smart, Open Directory, Snap, eBlast, etc., cuyas cifras son inferiores.

Ordenan sus recursos cronológicamente o geográficamente, por temas o por una combinación de estos. Pueden o no incluir un motor de búsqueda en su base de datos, son frecuentemente actualizados y coincidimos en que se han convertido en la consulta obligada de los profesionales de la información, tanto para temas generales como para búsquedas específicas.

Los directorios por lo general se organizan por categorías, entre otras opciones, algunos como Lycos, se auxilian de la tecnología para su actualización automática, en este caso, la de WiseWire. Sin embargo, un número significativo de editores en la elaboración de los listados, representa una buena señal en cuanto a que la calidad del directorio es superior.

Otra herramienta, Ask Jeeves (<http://www.askjeeves.com>), no está clasificado como directorio, pero funciona como ellos, es un servicio de respuestas que entrega 7 millones de ellas, su listado es compilado también por personas (30) y permite utilizar el lenguaje natural para ejecutar las preguntas.

### **Motor de búsqueda:**

- es una base de datos que indexa sitios Web
- es compilado por máquinas o robots (spiders, crawler, warm, wanderer)
- incorpora nuevos sitios automáticamente.

Los sitios son indexados de diferente modo por cada motor, pero en general lo hacen por título, primer encabezado, por URL, etc. y también difieren en los procedimientos que establecen para realizar las búsquedas.

"Scooter" el robot de AltaVista, por ejemplo, visita e indexa seis millones de páginas en un día, y se encuentra entre los cinco grandes y más populares motores de búsqueda, que procesan más de 75 millones de solicitudes de búsquedas diarias (HotBot, Northern Light, Excite, Infoseek, Lycos), no obstante su aceptación y gran popularidad, no puede afirmarse que estén exentos de dificultades. Ellos revisan constantemente la red, buscando nuevas páginas para incorporar a sus índices, pero tampoco abarcan a todo el Web cuando dan una respuesta a solicitudes a partir de palabras claves o frases, la respuesta dada está directamente relacionada con el contenido del índice del motor, de ahí que cada uno dará una respuesta diferente. Se recomienda para una búsqueda total, explorar más de una de estas herramientas.

Entre los aspectos a tener en cuenta a la hora de elegir uno de ellos, según Gil Castro, Virna E., están los siguientes:

- Facilidad de uso.
- Precisión de los resultados

Además:

- Tamaño de la base de datos, en número de URLs indexadas.
- Actualidad de la base de datos.
- Número de documentos proporcionados.
- Interfaz de búsqueda.
- Pantallas de ayuda en línea de cómo formular una consulta.
- Características de búsqueda.
- Lista de resultados, entre otros.

Adicionalmente, y siempre tratando de facilitarle el trabajo a quienes necesitan y buscan información, puede hablarse también de los **metabuscadores**, que aunque son pocos, se han consolidado, ellos buscan a través de varios motores (MB) y directorios de búsqueda (DB) de manera simultánea. Los resultados a menudo no se duplican y son re-rankeados según la relevancia, como Ej. Dogpile ([www.dogpile.com](http://www.dogpile.com)), que busca en catorce MB y DB diferentes, pero no elimina las duplicidades. También Mamma ([www.mamma.com](http://www.mamma.com)) el cual busca en siete MB, pero no duplica, y reordena según su algoritmo propio de relevancia, y MetaCrawler ([www.metaCrawler.com](http://www.metaCrawler.com)) que es uno de los primeros y máspreciados MB.

David Green, editor Jefe de Informed Business Services, considera que otro desarrollo interesante ha sido el de las búsquedas con lenguaje natural. Añade, que la mayoría de los MB sólo marcan exactamente las palabras solicitadas, con pequeñas consideraciones semánticas (ignoran las palabras relacionadas) o el uso de tesauros (algunos MB ofrecen un "más de esto" como opción, con resultados) y no consideran sintaxis (tales como palabras que aparecen en una oración). Los motores de lenguaje natural, tratan de resolver estos problemas, uno de ellos, Electric Monk ([www.electricmonk.com](http://www.electricmonk.com)) es apreciado por la relevancia de sus resultados, opera realizando un análisis sintáctico de los términos de búsqueda usando inteligencia artificial y algoritmos de lenguaje natural. Esto convierte los términos de búsqueda de los usuarios en una búsqueda compleja Booleana, que es entonces ejecutada en la base de datos de Altavista.

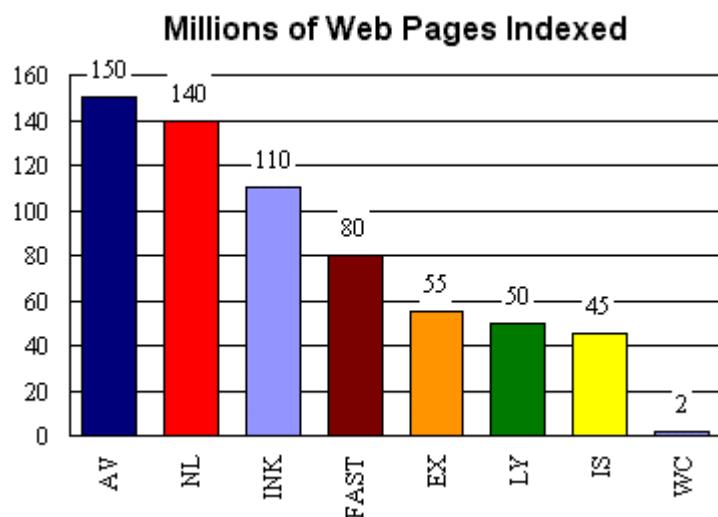
Veamos algunas de los gráficos comparativos compilados por Search Engine Watch (<http://searchenginewatch.com>), en los que se puede apreciar algunos elementos en la evolución y crecimiento de algunos de los motores de búsqueda

más grandes y conocidos en Internet, su cubrimiento del Web y las cifras o porcentajes según la cantidad de páginas indexadas.

En la gráfica que se presenta a continuación se hace una comparación entre siete de los más grandes motores de búsquedas.

**Gráfica comparativa actualizada del tamaño del índice de algunos motores:**

Los datos han sido reportados por cada uno de ellos el 1º de mayo de 1999

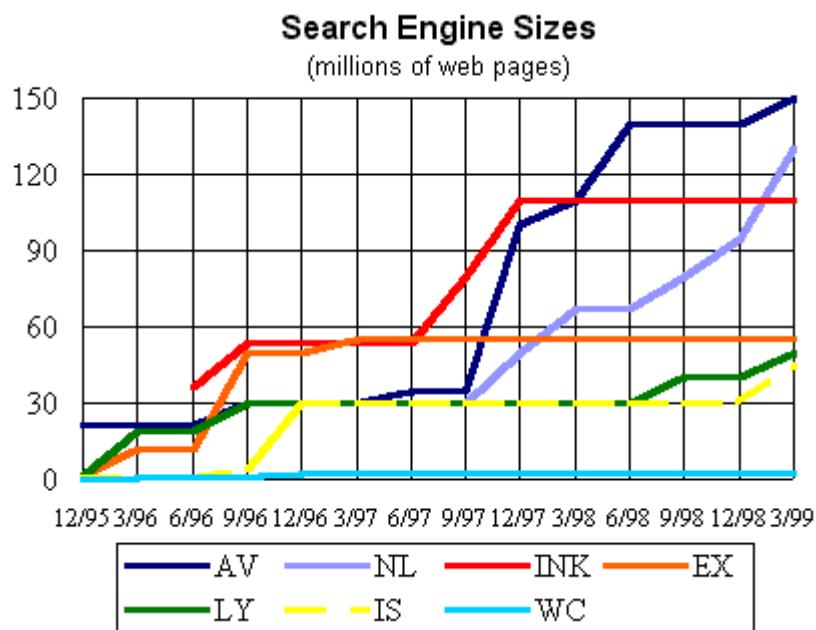


Leyenda: AV=AltaVista, NL=Northern Light, INK=Inktomi, FAST=FAST

EX=Excite, LY=Lycos, IS=Infoseek, WC=WebCrawler

AltaVista encabeza la gráfica, seguida por Northern Light, después Inktomi (potenciado por HotBot y MSN Search, entre otros) y le siguen los demás en orden descendente. Ellos son preferidos por bibliotecarios e investigadores, ya que por ser grandes índices dan mayores posibilidades de éxitos en la búsqueda, aunque este no sea un equivalente absoluto de buenos resultados.

**Gráfica del comportamiento en cuanto al tamaño de algunos motores de búsqueda en el período de diciembre de 1995 a marzo de 1999, refiriéndose a millones de páginas Web.**

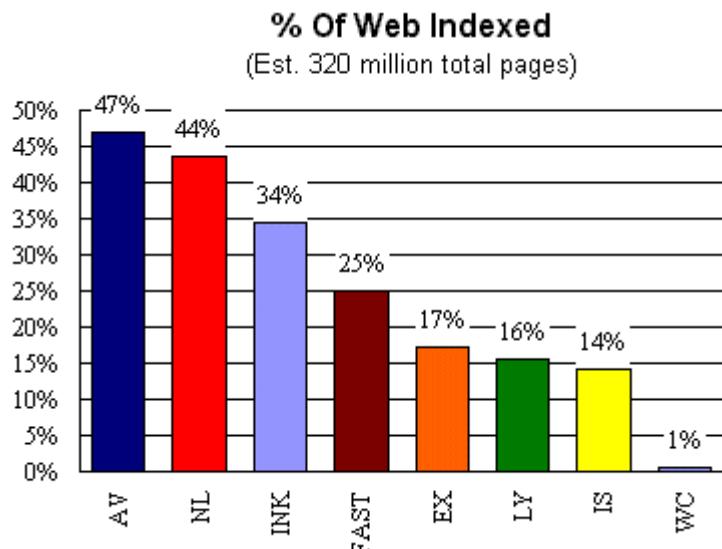


Cuando Altavista surgió, en diciembre de 1995, utilizaba un índice mucho mayor que ningún otro motor de búsqueda en esos tiempos. Por tanto, la competencia obligó a muchos de ellos a incrementar su tamaño a principios de 1996. De septiembre de 1996 a septiembre de 1997, ninguno de estos motores incrementó significativamente este indicador, a pesar que de hecho, el Web continuó su crecimiento ascendente. Así, de manera progresiva, el resto de los aquí referenciados, han incrementado paulatinamente la amplitud de sus índices, como lo demuestra este gráfico.

### Cobertura del Web

En abril de 1998, un estudio de la revista *Science* estimó que había alrededor de 320 millones de páginas indexadas en el Web. La gráfica que se muestra a continuación indica el porcentaje que cubre cada motor de búsqueda en correspondencia con el tamaño del motor de que se trate.

Forrester Research estimó en mayo de 1999 que existían alrededor de 550 millones de páginas indexadas. Este gráfico muestra también qué por ciento del Web es cubierto por cada motor de búsqueda, usando el estimado contra el tamaño de cada uno hasta ese momento.



Páginas relacionadas con Search Engine Watch en  
<http://searchenginewatch.com>

### **Tendencias actuales y evolución de los buscadores. La portalización**

Los principales motores de búsqueda ahora se convierten en un ciclo completo. La combinación de la focalización en la "portalización", junto a la competencia de nuevos motores de búsqueda con funcionalidad añadida ha forzado a los proveedores, a centrarse principalmente en los resultados de las búsquedas.

Tanto HotBot como AltaVista, han incorporado directorios. Altavista añadió lenguaje natural en octubre 98, y HotBot ingresó "Hit Direct", una nueva tecnología, en agosto. De hecho, estamos en el comienzo de un tiempo muy positivo para los que usamos el Web como herramienta de investigación (David Green, IBS.)

Según CMP Net Technology Encyclopaedia, en declaración a IWR (Internet World Review) en su número 45 de marzo del 99, **un portal está definido como un "supersitio" Web que provee de una variada cantidad de servicios de información a los consumidores, incluyendo búsquedas Web, noticias, directorios de páginas blancas y amarillas, grupos de discusión, ventas en línea y enlaces a otros sitios.**

José Antonio García Serrano, productor de Multimedia, plantea que la última novedad importante data de unos años atrás, y no tiene que ver con velocidades de acceso ni capacidades tecnológicas, y sí con la concepción de la Red como un inmenso y virgen campo de negocios. Este descubrimiento ha revolucionado uno de los aspectos que más ha ayudado al crecimiento de la Red, tanto en número de páginas e información publicada como en su desarrollo tecnológico. Y continúa expresando, que hasta hace poco los sitios que recibían un gran número de visitas

periódicas eran los que ofrecían los mejores servicios, buscadores, servicios de chat, hospedaje gratuito de páginas y Web e-mail. Desde hace aproximadamente un año ha surgido una nueva idea; ¿Por qué no concentrar todos estos servicios y reunir a todos los usuarios en un macro-sitio de servicios? De esta forma ha nacido un nuevo concepto de sitio: **el portal de Internet, una página que pretende ser la entrada a la Red de todo internauta.**

Explica además, que el gran reto de un portal para estar entre los grandes y conseguir captar un gran número de usuarios (lo que se traduce en publicidad y ventas por comercio electrónico) es ser atractivo y responder a las necesidades del usuario, de forma que siempre vuelva a él cuando está perdido, cuando necesita información, una cuenta e-mail, un diccionario, etc. Esto no es una tarea fácil, ya que los demás sitios siempre están "a un clic" y nada obliga a volver. Un portal que pretenda obtener una posición privilegiada debe disponer de un sistema de búsqueda, servicio de chat y foros, opcionalmente e-mail y hospedaje gratuito, comercio electrónico y un servicio de noticias personalizado.

Portales destacados en Internet	
<u><a href="http://www.yahoo.es/">Yahoo!</a></u>	<a href="http://www.yahoo.es/">http://www.yahoo.es/</a>
<u><a href="http://www.excite.com/">Excite</a></u>	<a href="http://www.excite.com/">http://www.excite.com/</a>
<u><a href="http://home.microsoft.com/Default.asp">MSN</a></u>	<a href="http://home.microsoft.com/Default.asp">http://home.microsoft.com/Default.asp</a>
<u><a href="http://www-es.lycos.com/">Lycos</a></u>	<a href="http://www-es.lycos.com/">http://www-es.lycos.com/</a>
<u><a href="http://www2.telepolis.com/">Telépolis</a></u>	<a href="http://www2.telepolis.com/">http://www2.telepolis.com/</a>
<u><a href="http://www.ciudadfutura.com/">Ciudad Futura</a></u>	<a href="http://www.ciudadfutura.com/">http://www.ciudadfutura.com/</a>
<u><a href="http://www.ole.es/">Olé</a></u>	<a href="http://www.ole.es/">http://www.ole.es/</a>

Para conseguir esta concentración se ha desatado una fiebre de compras y absorciones, iniciada en EE.UU., donde empresas de servicios han sido absorbidas por buscadores y estos a su vez por empresas de telecomunicaciones. Nos podemos hacer una idea de la importancia que tienen estos nuevos lugares de entrada en Internet por los valores de compra que alcanzan en el mercado: US\$6.700 Excite, US\$3.500 Geocities o los 3.000 millones de pts. en que se valora el español Olé, continúa diciendo el Sr. García Serrano.

Y acota: "la nueva situación exigirá creatividad para llevar los portales más allá de la simple concentración de servicios existentes que, indudablemente atrae audiencias existentes, pero no las aumenta. Además el continuo desarrollo de las telecomunicaciones abrirá paso a nuevas posibilidades, hoy ya tenemos fibra

óptica y otras formas de transmisión de la señal "a la puerta de casa", mañana tendremos esa mayor capacidad de transmisión disponible en nuestro ordenador y será necesario que los portales innoven en sus servicios para aprovechar estas capacidades."

### **Factores claves del éxito de un portal. Otro punto de vista.**

David Boronat, consultor de la agencia Ogilvy Interactive (<http://www.ogilvyinteractive.es>), ha elaborado un interesante informe sobre la situación de los portales en España, su presente y su futuro.

Entre los datos proporcionados por el informe, realizado en base a encuestas a través del Web, están los factores clave que llevan a un portal al éxito. Estos factores han sido analizados por un grupo de expertos, y destacaron los siguientes, por orden de importancia:

- 5.2.1.1. Capacidad financiera
- 5.2.1.2. Notoriedad y Marca
- 5.2.1.3. Alianzas estratégicas
- 5.2.1.4. Promoción on-line y off-line
- 5.2.1.5. Nivel de contenidos
- 5.2.1.6. Servicios de valor añadido
- 5.2.1.7. Sistema de búsqueda
- 5.2.1.8. Penetración en el mercado (visitas)
- 5.2.1.9. Sentimiento de comunidad
- 5.2.1.10. Orientación a la localidad
- 5.2.1.11. Conocimiento del mercado
- 5.2.1.12. Infraestructura tecnológica y personal
- 5.2.1.13. Grado de especialización
- 5.2.1.14. Antigüedad y experiencia
- 5.2.1.15. Usabilidad
- 5.2.1.16. Funcionalidad
- 5.2.1.17. Acceso al mercado
- 5.2.1.18. Nacido del propio medio
- 5.2.1.19. Proyecto internacional
- 5.2.1.20. Globalista - Generalista
- 5.2.1.21. Pertenencia al mundo off-line

Mucho más puede decirse acerca de los portales, su clasificación, tendencias y posibilidades de permanencia como tales, pero no es el objetivo de este trabajo, se ha hecho necesario abordarlos someramente, por su relación directa con las búsquedas en el Web y la influencia positiva que ellos tienen en otros servicios dirigidos al usuario.

## **Sobre la importancia de una adecuada y efectiva estrategia de búsqueda para obtener buenos resultados.**

Son diversas las opiniones revisadas al respecto, y la coincidencia entre ellas ha sido muy didácticamente expresada y resumida por Gil Castro, Vilma E., de la Biblioteca ITESM Campus Sinaloa, en su Compendio sobre Búsquedas, vigente aún en la actualidad. En él se puntualiza que:

Independientemente del motor de búsqueda utilizado, el desarrollo de una efectiva estrategia de búsqueda es fundamental si uno espera obtener buenos resultados. Una estrategia de búsqueda genérica sigue los siguientes pasos:

1. Formulación de la pregunta de búsqueda y su alcance.
2. Identificación de conceptos importantes dentro de la pregunta.
3. Identificación de términos de búsqueda para describir estos conceptos.
4. Consideraciones de sinónimos y variantes de estos términos
5. Preparación de una búsqueda lógica.

Esta estrategia puede ser utilizada en la búsqueda en cualquier herramienta de información electrónica, incluyendo catálogos de biblioteca y bases de datos en CD-ROM. Añade, que una bien planeada estrategia de búsqueda es de especial importancia cuando se toma en consideración que la World Wide Web es una de las más grandes y amorfas del mundo. Un factor que subraya la necesidad de una estrategia efectiva es el hecho de que la mayoría de los motores de búsqueda indexan cada palabra del documento. Este método tiende a incrementar el número de resultados obtenidos de la búsqueda, al mismo tiempo que deteriora la relevancia de estos resultados, por el incremento de palabras parecidas encontradas en un contexto inapropiado. Cuando se selecciona un motor de búsqueda, un factor a considerar es si éste permite al buscador especificar cual parte o partes del documento debe buscar (en URL, el título, primer encabezado...) o si simplemente el motor buscará en el documento completo.

## **Consideraciones para la selección de buscadores de acuerdo con el interés y complejidad de la búsqueda.**

Gil Castro plantea, que no existe un sistema de búsqueda que sea realmente "el mejor". Ninguno de los buscadores puede incluir todas las bases de datos de Internet. Y como el proceso de indexado y evaluación de los sitios es también diferente, ningún sitio puede clamar de tener todo en este punto. Para una ampliación y orientación a la hora de escoger uno de ellos, debe verse el Anexo No1. Se recomienda también ver las características de algunos de los buscadores aquí mencionados de manera más particular en: Compendio sobre Búsquedas en la dirección

<http://campus.sin.itesm.mx/~biblio/netcurso.html#estrategia>

Adicionalmente, ya conocemos los beneficios de los bookmarks como facilitadores para la navegación, al tener compiladas todas las direcciones de sitios de nuestro interés y un enlace directo a ellos sin teclear una palabra, sin embargo, recientemente ha surgido una nueva herramienta que potencia aún más la navegación, se trata de los Bookmarklets, con su servicio More Info About. Son herramientas sencillas que amplían las posibilidades de búsqueda y de navegar de los conocidos Browsers Netscape e Internet Explorer, gratis además. Permiten modificar la forma de ver algo en una página Web, extraer datos desde una de ellas, buscar con mayor rapidez, de una manera que no es posible hacerlo con un motor de búsqueda, y navegar por nuevas vías. Trabaja en plataformas Windows, Macintosh y Unix, y no es necesario descargar o instalar software alguno para usarlo. Es de esperar que en breve alcance buena popularidad, porque tiene la ventaja de dejarnos dialogar en la búsqueda a partir de recursos ya clasificados, digamos por noticias, directorios, diccionarios, etc. sin tener que salir a buscar una dirección de algún buscador, colocando este sitio en nuestra barra de herramientas, sólo hay que abrirlo para decidir por cual vía tomamos para realizar la búsqueda.

Su lanzamiento inicial fue en diciembre de 1998. En Bookmarklets What's New <http://www.bookmarklets.com/tools/new.html> se puede localizar información muy completa y actualizada sobre esta herramienta.

## **Conclusiones**

Luego de recorrer y analizar los datos que han sido compilados a partir de intensas búsquedas en la Red, y de nuestra propia experiencia en el uso de la misma, y confrontar la opinión de diferentes autores sobre los buscadores, las búsquedas y el modo de hacerlas cada vez mejor y más rápida y eficientemente, podemos arribar a las siguientes conclusiones:

- 1.- Los buscadores en el Web constituyen una valiosa herramienta para obtener información, insustituibles, si se logra escoger el más adecuado, según el tipo de información que necesitamos, y si se tiene en cuenta tanto a directorios, a motores de búsqueda, como a los metabuscadores.
- 2.- El factor humano en la confección de los listados predefinidos de los directorios, da ventajas cualitativas en los resultados de búsquedas que puedan obtenerse por encima de los motores de búsqueda, que utilizan robots, cada uno con sus especificidades.
- 3.- El más potente de los directorios conocidos hasta el momento, por su tamaño, cobertura en el Web, velocidad de respuesta y desarrollo en cuanto a servicios de valor agregado es Yahoo!, lo cual no significa que sea el mejor de todos, cada uno tiene intereses diferentes y ninguno cubre a totalmente el Web.

4.- Altavista es, de los motores de búsqueda, además de muy popular, efectivo y eficiente en la ejecución de búsquedas, el que mayor cantidad de páginas tiene indexadas, con un crecimiento sostenido desde su surgimiento, y permite hacer búsquedas complejas.

5.- Existe una relación muy directa entre las dimensiones del buscador y sus posibilidades de devolver buenos resultados de búsqueda, si se ha logrado elaborar una adecuada estrategia al efecto.

6.- Como resultado de las actuales tendencias en el desarrollo de servicios de valor agregado, diversos sitios hasta hace poco muy conocidos sólo como buscadores en el Web, han evolucionado hacia una fase diferente y superior: los portales de Internet, manteniendo sus objetivos iniciales en apoyo a las búsquedas, pero ampliando las posibilidades a los usuarios, al contar con servicios cada vez más personalizados y de mayor diversidad, e igualmente fortalecer su influencia en el empleo del comercio electrónico a través de la Red.

7.- El desarrollo sostenido de nuevos productos y servicios en la Red para aumentar las posibilidades de acceso a la información que necesita el usuario, resulta de gran beneficio para todos, pues sistemáticamente surgen nuevas modalidades y variedades de ellos, que si no sustituyen a otros, en muchos casos los potencian.

## Bibliografía

Abilock, Debbie. "MLA Interactive Citation Format for a World Wide Web Page." *Nueva Library Help*. 22 agosto 1996.  
<<http://www.nueva.pvt.k12.ca.us/~debbie/library/research/cit/mla/citwww.html>> (17 junio 1999).

Barlow, Linda. "Search Engine Details." The Spider's Apprentice. 22 May 1997.  
<<http://www.monash.com/spidap3.html>> (10 junio 1999).

"Bookmarklets Home Page" <<http://www.bookmarklets.com/>> (16 junio 1999)

Cohen, Laura. Searching the Internet: Recommended Sites and Search Techniques. 23 May 1997. <<http://www.albany.edu/library/internet/search.html>> (21 mayo 1999).

Gil Castro, Virna E."Búsquedas en Internet y Evaluación de la información." *Internet, un medio de consulta.* julio 1997.  
<<http://campus.sin.itesm.mx/~biblio/netcurso.html#estrategia>>. (junio 13 1999)

Green, David. "Search Insider." *Information World Review November, 1998* Pag 46 - 47.

<<http://www.bookmarklets.com/tools/new.html>> *Bookmarklets What's New* (15 junio 1999)

<<http://www.bookmarklets.com/moreinfo.phtml>> *More Info About.* (16 junio 1999)

iWorld: Noticias. *Archivo.* [http://208.164.21.2/iworld/archivo\\_not.asp](http://208.164.21.2/iworld/archivo_not.asp) (13 mayo 1999)

Murray, Janeth. "Comparing Search Engines". Searching the Net: An Online Internet Institute Project. Wilson High School. Portland. Or. - No date-

<<http://arlo.wilsonhs.k12.or.us/compare.html>> (11 Junio 1999).

Notess, Greg R. "Comparing Net Directories". Database 20.1 Feb 1997

<<http://www.onlineinc.com/database/FebDB97/nets2.html>> (28 mayo 1999).

Sullivan, Danny. " Tips and more about Search Engines" *Search Engine Watch: News*, <<http://searchenginewatch.com/reports/index.html>>. (junio 16 1999)

Sullivan Danny. *Direct Hit Debuts Personalized Search* \_2 febrero 1999

<<http://searchenginewatch.com/sereport/9902-directhit.html>>, (14 mayo 1999)

## Anexos

### Sugerencia de alternativas de buscadores para efectuar búsquedas complejas.

**Si desea:**

1. Información limitada específica.
2. Información científica de respaldo.
3. Información sobre un nombre propio –lugar, persona-
4. Información obscura o desconocida.
5. Información de

**Utilice:**

Alta Vista <http://altavista.digital.com>

actualidad.

1. Un número pequeño de resultados relevantes con ligas a otras páginas similares. Excite <http://www.excite.com>
2. Una idea de qué hay en cada documento antes de ir a cada página.
  1. Palabras correctas de una cita y/o autor. Open Text <http://www.opentext.com>
  2. Buscar frases completas.
    1. Información específica. Hot Bot <http://www.hotbot.com>
    2. Información de un evento ocurrido en una fecha que conoces.
    3. Información sobre un lenguaje de programación.
    4. Información sobre un nombre propio –lugar, persona–
    5. Información de páginas con un dominio de Internet.
    6. Buscar archivos de sonidos o imágenes.
    7. No buscar frases completas.

8. Información geográfica.
1. Describir tu tema Infoseek <http://www.infoseek.com>  
con lenguaje natural. Ultraseek
  2. Buscar nombres de personas, cuyas palabras puedan confundirse con temas distintos. – Sensitividad y operadores de proximidad-  
3. Información sobre stocks. Símbolo de una compañía que cotiza en bolsa.
- 
1. Buscar imágenes, sonidos, tipos específicos de archivos de información. Lycos <http://www.lycos.com>
  2. Información sobre temas comunes.
- 
1. Encontrar páginas populares o de mucha actualidad y/o de temas muy comentados. WebCrawler <http://www.webcrawler.com>
  2. Hacer una búsqueda simple.
- 
1. Información común o conocida –no términos WWWorm <http://www.cs.colorado.edu/home/mcbryan/WWW.html>

oscuros- .

2. Encontrar páginas con imágenes GIF.
  3. Información en relación con educación.
- 
1. Tener una idea Yahoo <http://www.yahoo.com> del tema.
  2. Información jerarquizada por temas.
  3. Navegar por Internet y no sabes por donde empezar.
  4. Información sobre compañías.
  5. Información sobre sitios comerciales.
  6. Información sobre patentes.
  7. Acceso a la base de datos temática más grande de Web.
  8. Un medio de referencia recomendado por especialistas de bibliotecas.
- 
1. Un directorio Lycos Top 5% <http://point.lycos.com/categories> temático.
  2. Descripción de A2Z <http://a2z.lycos.com> las páginas proporcionadas y calificación de evaluación.
  3. Búsqueda por 16 categorías principales y/o subcategoría.

4. Un medio de referencia recomendado por especialistas de bibliotecas.

1. Un directorio temático con motor de búsqueda. Magellan. <http://www.mckinley.com>

2. Búsqueda a través de 26 categorías.

3. Descripción de las páginas con su evaluación.

4. Bases de datos de patentes.

5. Un medio de referencia recomendado por especialistas de bibliotecas.

1. Un directorio temático con un motor de búsqueda. Excite Reviews <http://www.excite.com/Reviews>

2. Búsqueda a través de 16 categorías.

1. Un directorio altamente selectivo.

2. Cobertura de áreas académicas.

3. Información evaluada por especialistas en el tema.

4. Información en su mayoría en texto sin

Argus Clearinghouse <http://www.clearinghouse.net>

vínculos.

5. Información recomendada por especialistas.
  1. Comparación de resultados entre varios buscadores. SavvySearch <http://www.cs.colostate.edu/~dreiling/smatform.html>
  2. Buscar en varios buscadores al mismo tiempo.
1. Comparación de resultados entre varios buscadores. Metacrawler <http://metacrawler.cs.washington.edu:8080>
2. Buscar en varios buscadores al mismo tiempo.
3. Páginas de una región geográfica o dominio de Internet.

## Búsqueda y Recuperación de Información en la Internet

José A. Mañas <jmanas@dit.upm.es>

Dpto. de Ingeniería de Sistemas Telemáticos

E.T.S. de Ingenieros de Telecomunicación

Universidad Politécnica de Madrid

30 de junio, 1994

### **Resumen**

Internet permite el acceso a cantidades ingentes de información, acceso que se logra en dos fases: localización y recuperación; de las cuales la primera es la más difícil, con diferencia. En este artículo presentamos una serie de herramientas, casi todas muy recientes, que ayudan a moverse por la red, a organizar su contenido y a descubrir lo que se busca.

### **Introducción**

La Internet es un conglomerado ingente de recursos. Se calculan en 10-20 millones las personas accesibles, y del orden de billones los bytes de documentación en general y programas que se pueden recoger de la red. En semejante abundancia es necesario disponer de medios eficaces para localizar lo que se necesita y traerlo al ordenador local.

Por razones históricas, la Internet se ha venido desarrollando de forma abierta y voluntarista. Esta es la principal razón de su éxito: al no ser de nadie, todo el mundo la considera un poco suya, y aporta su granito de arena. Pero al tiempo es su debilidad: con todo el mundo aportando por su cuenta el orden es escaso, y la calidad no es homogénea. Esta situación resulta chocante en un desarrollo tecnológico de finales del siglo XX; pero en realidad es lo habitual en obras humanas tan ambiciosas.

La Internet se parece mucho a una gran ciudad, donde las redes de comunicaciones proporcionan a modo de calles y accesos para llegar a los servicios que se ofrecen a los visitantes. Hay barrios interesantes y barrios peligrosos; hay tiendas de andar por casa, supermercados y tiendas de lujo; hay zonas que vale la pena visitar, y otras que no; hay ciudades universitarias y parques tecnológicos, etc. Como en toda ciudad un poco grande (y la Internet es muy grande), se necesitan planos y guías para saber a dónde ir y dónde encontrar lo que se necesita.

A diferencia de una ciudad, la Internet se puede recorrer cómodamente desde casa, moviendo poco más que los dedos sobre las teclas. E igualmente las guías y planos de Internet son accesibles como un servicio más. "No me des un pez!, enséñame a pescar!" dice el viejo proverbio chino, y así intentaré hacerlo a

continuación: no voy a contar dónde están los servicios finales, sino sólo dónde están las guías, qué información proporcionan, y un poquitín de cómo funcionan.

La Internet permite que cualquier par de ordenadores conectados a ella puedan comunicarse entre sí de una forma que parece directa; es decir, que es capaz de ocultar en buena medida la forma física de llegar de uno a otro. El límite lo impone la calidad de los enlaces de comunicaciones que realmente utilicemos.

Los ordenadores se comunican por medio de **protocolos** que son lenguajes monotemáticos que permiten un cierto nivel de diálogo e incluso de negociación entre equipos. En Internet hay multitud de protocolos, cada uno especializado en su pequeña área de aplicación. Pero hay una cosa común a prácticamente todos los protocolos que se utilizan en Internet: el modelo **cliente-servidor**. Si un ordenador ofrece algo al exterior, se dice que es un centro servidor. Técnicamente, un servidor es un programa que está permanentemente escuchando lo que se dice por Internet y cuando oye algo en su idioma (en su protocolo) se activa y contesta. Las comunicaciones las empiezan los clientes que, a petición del usuario final, se ponen en comunicación con el servidor para lograr un cierto servicio distribuido entre ambos.

Es fácil confundir un protocolo con una aplicación que lo utiliza. A ello ayudan los programadores, que le suelen llamar a todo con el mismo nombre. Para el usuario final, que no se supone un experto en programas y protocolos de comunicaciones, todo esto debe dar un poco igual. Lo que hay que saber es qué servicios hay para poder elegir el que nos conviene, y saber un poquitín de sus características técnicas para fundamentar la decisión.

El resto del artículo se divide en dos grandes bloques: primero presentaré herramientas para localizar programas y documentos en general, y en segundo lugar describiré servicios de información sobre personas. A lo largo del texto utilizaré la notación URL (identificación única de recursos en Internet) que se describe en el Apéndice A.

## Localización de Programas y Documentos

Desde tiempos inmemoriales Internet proporciona un protocolo/herramienta denominado FTP [14] que permite transferir ficheros entre ordenadores. Se basa en un servidor de ficheros y un cliente de acceso que comunican los directorios de dos ordenadores permitiendo moverse por ellos e intercambiar ficheros.

FTP ha sido tradicionalmente el protocolo estrella de la Internet en términos de volumen de datos transferidos. Este papel se ha visto relativizado con la proliferación de aplicaciones multimedia que provocan trasiegos mucho mayores.

Conociendo el nombre del ordenador, el directorio y el nombre del fichero, es elemental recuperarlo. A veces se necesita una cuenta y palabra de paso, si bien

la información pública suele estar bajo una pseudo cuenta anónima y como palabra de paso se utiliza la dirección de correo electrónico del peticionario (por si hubiera que contactarle en caso de problemas).

### **Archie**

Todo es muy fácil con FTP en cuanto conocemos el ordenador, el directorio y el nombre del fichero; si alguien nos lo dice, perfecto; pero si no, hay que buscarlo. Para ello se desarrolló Archie [\[5\]](#) [\[6\]](#) en la Universidad de McGill (Canadá).

Esta es una herramienta que recopila información (directorios y ficheros) de una larga serie de servidores FTP anónimos, la indexa, y permite búsquedas sobre ella. La tarea de acceder a servidores FTP es lenta y actualmente lleva más de un mes completar un ciclo de consultas. Originalmente había un sólo centro que hacía todas las consultas. Actualmente hay una red de centros que se reparten la tarea de consultar y luego intercambian información entre ellos.

Archie mantiene también un enorme índice que correlaciona nombres de ficheros con breves descripciones de su utilidad o contenido (1 línea). Esto permite consultas por contenido o función para averiguar el nombre del fichero preciso.

Los servidores archie suelen proporcionar una cuenta anónima desde la que se puede acceder esta información. Además, cualquier ordenador conectado a Internet puede instalar su cliente que permite elaborar las consultas localmente y luego sigue un protocolo específico para consultar al servidor más cercano.

Actualmente, archie mantiene información de más de un millar de servidores anónimos, indexa más de 4.000 paquetes y más de 2 millones de documentos. Se realizan unas 50.000 consultas diarias.

Archie es tan efectivo que este artículo no necesita dar indicaciones de dónde se puede encontrar cada herramienta de las que se citan: "!consulte a archie!".

### **Wide Area Information Servers**

Los bibliotecarios son unos profesionales muy interesados en el uso de la Internet como medio de acceso a las bibliotecas. Por eso llevan años trabajando en un protocolo que permita transacciones entre los archivos informatizados de las diferentes bibliotecas. Siguiendo el modelo cliente--servidor, el cliente de una biblioteca puede ponerse en contacto con el servidor de otra para recabar información, básicamente índices de fichas de existencias. De esto entiende el protocolo [Z.39-50 \[12\]](#) desarrollado por NISO y en vías de aceptación mundial por ISO (10163).

Brewster Kahle, mientras trabajaba en Thinking Machines, Inc., y en colaboración con Dow Jones, Apple y KPMG Peat Marwick, desarrolló una herramienta, [WAIS](#)

[9], basada en dicho protocolo. Actualmente, B. Kahle ha montado su propia empresa en base a esta herramienta.

WAIS permite indexar localmente la información que tengamos y acceder a ella bien localmente (uso privado) o remotamente (uso público, que puede ser gratuito o de pago). La herramienta admite consultas en lenguaje natural, identifica posibles documentos que respondan a la pregunta, e incluso permite recuperar el documento en sí.

Es esencial la filosofía de que la inteligencia está en el servidor. Se quiere que los clientes sean muy sencillos conceptualmente (interfaz con el usuario y punto) con lo que pueden difundirse ampliamente y de forma gratuita. Lo sofisticado es el servidor que ante una pregunta en lenguaje natural busca inteligentemente por los índices y devuelve referencias a documentos que pueden casar. El servidor debe admitir casi cualquier pregunta, y debe responder sensatamente, no sólo con referencias a documentos que casan con la pregunta, sino que además debe ser capaz de ordenar estos documentos. Toda ordenación es heurística y cada servidor puede aplicar diferentes heurísticos, es más, puede ir aprendiendo con el tiempo y contestar cada día con más fundamento. El mismo cliente puede hablar con muchos servidores, tanto si son simples y sus respuestas tienen un valor muy relativo, como con servidores muy elaborados y costosos. Estabilizado el protocolo, ahora el negocio está tanto en la información como en la inteligencia de los servidores.

La ordenación de las respuestas se denomina "*realimentación significativa*" y da pie a una interesante área de aplicación de las técnicas de inteligencia artificial y lógica borrosa. El cliente puede formular la pregunta como una frase normal. El servidor elimina signos de puntuación, artículos, conjunciones, palabras muy comunes (que en la práctica aparecen en todos los documentos) y hace algún análisis de plurales, tiempos de los verbos, etc. Al final, se queda con una serie de palabras que considera significativas, busca en los índices, y ordena los documentos que contienen dichas palabras para responder al cliente. La ordenación responde a heurísticos que pueden ser tan simples como el número de veces que aparece cada palabra, si aparece en el título o perdida en el contenido, si las palabras aparecen juntas, cercanas o dispersas, etc. Por último, WAIS puede tomar un documento como referencia: el cliente pide "búscame más que traten de lo mismo que éste". WAIS extrae del documento aquellas palabras que le parecen significativas (no son términos comunes, aparecen muchas veces) y aplica estos datos para lanzar una nueva búsqueda.

Por ejemplo, si preguntamos acerca de *individuals* y *privacy*. Una vez localizado un artículo interesante, hemos vuelto a lanzar una segunda consulta buscando artículos similares. El resultado de esta última consulta aparece en la figura.

El resultado global es que el usuario puede partir de una pregunta muy general e ir refinando su búsqueda de forma interactiva. Lo importante es que bases de datos

muy complejas puedan ser utilizadas por gente de la calle sin requerir una formación específica en bases de datos.

### **Navegación por la Internet**

Recuperando el símil de la Internet con una gran ciudad, antiguamente la única forma de llegar a un ordenador era plantarse en él como a través de una quinta dimensión. Esto es excelente cuando uno sabe a dónde va; pero a menudo se necesita más. Cuando uno va de compras, una forma de encontrar lo que se busca es callejear por el barrio adecuado mirando en las diferentes tiendas. En Internet a esto se le llama **navegar**.

### **Gopher**

La necesidad de navegar se convirtió en imperiosa en la Universidad de Minnesota (1990) que dispone de un amplio campus y multitud de recursos informáticos: científicos, académicos, administrativos, culturales, etc. Se estaba convirtiendo en un lío el mantener un directorio en papel del nombre de cada ordenador, la información que contenía, dónde estaba, etc., y así se inventaron el gopher [11]. El nombre es el de un ratoncito habitual en aquellas praderas que es capaz de encontrar lo que busca a través de un laberinto de canales.

La idea es sencilla: convertimos cada directorio público en un menú; de forma que un usuario disponga de una interfaz sencilla para elegir a dónde quiere ir. Si elige un fichero, se le da; si elige otro directorio, aparece un nuevo menú y así sucesivamente. En la práctica los nombres de los directorios y ficheros suelen ser crípticos (en un PC se limita a 8 caracteres, en muchas máquinas Unix a 15 caracteres, etc.). Así gopher propone un mecanismo complementario para asociar a cada objeto una breve frase descriptiva (normalmente 1 línea) para hacer menús cómodos e inteligibles.

El siguiente paso de gopher son los directorios virtuales o menús cuyos componentes no son necesariamente los ficheros de un ordenador, sino que puede saltar a los de otro ordenador, todo esto oculto bajo la frase descriptiva que el usuario puede seguir utilizando para saltar de ordenador en ordenador sin darse cuenta (bueno, es un decir, la velocidad de acceso puede cambiar sustancialmente; pero la interfaz funcional es uniforme).

Con esta filosofía, todos los ordenadores que instalen un servidor gopher quedan integrados en el entramado de calles y avenidas. Está en la mano de los administradores (bibliotecarios) el ir construyendo menús virtuales que ordenen el planeta por servicios, tarea que avanza a pasos agigantados.

El protocolo gopher se ha diseñado de forma abierta, tal que permita extensiones fáciles del tipo de documentos que se traen. Los clientes gopher disponen de una tabla de configuración similar al mecanismo de *asociación* de MS-Windows. En MS-Windows se asocian las extensiones en el nombre de ficheros con una aplicación capaz de manejarlo. En gopher se asocia un tipo de fichero con una

aplicación. El tipo lo decide el servidor (probablemente en base a la extensión del nombre) y la asociación la realiza el cliente. Con este simple mecanismo es posible introducir información multimedia: imágenes y sonido. Nada impide asociar ficheros para hojas de cálculo, bases de datos, etc.

Incluso es posible mantener el mismo material en diferentes formatos. Por ejemplo, en castellano, inglés, ASCII, PostScript, Word, etc. El cliente avisa al usuario de la variedad existente, y éste puede elegir la opción que más le convenga.

Gopher se complementa con una escueta capacidad de dialogo. Ciertos enlaces se califican como indexados, lo que se interpreta como que el cliente gopher debe pedirle al usuario una relación de palabras que quiere utilizar en la búsqueda. Con esta información, el cliente se pone en contacto con el servidor, le pasa las palabras de búsqueda, recoge la relación de documentos que localiza el servidor (ordenados por algún heuristicó que permita una realimentación significativa) y los presenta como un nuevo menú en el que el usuario puede elegir algún documento, etc.

Globalmente, todo el gopher es bastante sencillo, el protocolo, el servidor y, sobre todo, el cliente, por lo que existen multitud de versiones para todo tipo de modelos de ordenadores.

Cuando los proveedores detectaron este invento, se les hizo la luz para anunciar sus productos en la red. Es obvio cómo organizar sus preciosos catálogos de productos en forma de directorios de clasificación y ficheros de descripción, precios, disponibilidad, etc. Y es obvio que aparece un negocio adicional que es de creación de servicios de páginas amarillas que agrupen proveedores.

## **World Wide Web**

Mientras gopher crecía y se desarrollaba en las praderas de Minnesota, los europeos estábamos incubando un huevo similar; pero mucho más potente. Tim Berners-Lee del CERN (el laboratorio europeo de física nuclear en Ginebra) desarrollaba la "World Wide Web" [2], WWW para simplificar. Comparte muchas ideas con gopher; pero aquí se basa en ficheros como elementos de menú para navegar. Esto se llama **hipertexto** y consiste en hacer activos los ficheros. Ciertas palabras, frases o incluso dibujos son algo más que una cosa para observar: si uno los activa, el sistema nos lleva a otro documento bien en el mismo ordenador, bien en otro ordenador. En esta filosofía, los directorios se transforman en ficheros activos, y todo queda homogéneo.

WWW es, entre otras cosas, un procesador de textos. El que prepara un documento para hipernavegar lo hace rellenando el documento de claves para ordenar párrafos, listas, negritas, etc., como en los peores procesadores de texto de la antigüedad (van apareciendo poco a poco procesadores de texto interactivos tipo Word; pero aún no está el mercado muy desarrollado). El caso es que el que

prepara el documento no sabe si el lector dispondrá de una preciosa pantalla de colores y alta resolución o un simple terminal vt100. La idea es que el cliente haga el procesado del texto para adecuarlo a las características del terminal del cliente. Igualmente hay que decidir de qué forma el usuario identifica un enlace hipertexto para seguirlo; si tiene ratón es bastante obvio, pero en terminales baratos hay que ponerle cierta imaginación adicional.

La gran diferencia de WWW frente a gopher está en basarse en documentos activos. El documento bien formateado, con figuras y explicaciones es la forma natural de comunicarse con el lector. En gopher un documento está muerto: para seguir navegando hay que salir de él y regresar a los menús de directorios. En WWW se navega desde el cuerpo de los documentos.

Al igual que el gopher WWW es capaz de tratar información multimedia a base a un mecanismo de asociación de tipos de ficheros con aplicaciones que lo manejan. Para ello WWW se basa en una norma ampliamente aceptada: [MIME](#) [3], que permite intercambiar objetos de varios tipos. Sin embargo, introduce una importante diferencia con respecto a gopher y es que el protocolo que usa WWW permite una cierta negociación entre el cliente y el servidor. Si un cliente no está capacitado para tratar un cierto tipo de información, el servidor no se la envía. Es más, el cliente puede informar al servidor de sus características (por ejemplo la resolución de la pantalla) y el cliente puede decidir no enviar una información si ello supusiera una degradación considerable de calidad.

Parece obvio que WWW atraiga mucho más aún a los proveedores que pueden poner en la red sus hermosos catálogos en toda su plenitud, sin esclavitud a los menús. Pueden intercalar figuras, índices, y de todo, y hacerlo todo activo para ir dirigiendo al visitante hasta que encuentre algo que satisfaga sus necesidades.

WWW difumina la frontera entre la información personal y la pública. Un usuario de Internet puede perfectamente organizar su puesto de trabajo en base a documentos WWW que enlazan en los puntos oportunos con otros recursos Internet. Toda Internet es, virtualmente, mi puesto de trabajo. Esta posibilidad de salto sin solución de continuidad se utiliza constantemente para organizar directorios globales, o presentaciones (con demostraciones reales incrustadas), o cursos de autoformación, o lo que se quiere dentro de un amplio abanico de posibilidades.

## **Localización de Personas**

Una de las preguntas más frecuentes de los recién llegados a Internet es "¿cómo encuentro a una persona que se llama así y está en tal sitio trabajando?". Aún siendo una pregunta elemental, sigue poniendo nerviosos a los *gurús* de la red, pues la tarea de encontrar a alguien no es trivial, exige dominar varias herramientas y completar con imaginación la distancia entre unas coordenadas

imprecisas (que son las que suele proporcionar el que busca) y la precisión de un ordenador.

Estos servicios se suelen denominar "páginas blancas" o, simplemente, "directorios" [13]. Hay una serie de ellos, muy diferentes entre sí y poco o nada interconectados. Parece obvio que en un mundo heterogéneo los servidores de páginas blancas serán heterogéneos; pero también debería ser obvio que los clientes deberían ocultar esta heterogeneidad bajo una interfaz única. Esto dista de ser tarea fácil, pues hay que combinar varios protocolos y, además, cada servidor tiene un contenido diferente y no necesariamente homologable. En definitiva, estamos ante un problema de bases de datos heterogéneas y distribuidas, para el que no se conoce una solución definitiva.

Algunos de los servicios actualmente disponibles permiten enlazar centros servidores y navegar. Otros, la mayoría, funcionan aisladamente. Si el usuario logra saber a quién debe preguntar, el servidor estará encantado de responderle con su información local. Y punto.

## **X.500**

CCITT intentó poner orden con la normalización de un servicio de directorio denominado X.500 [4]. Pese a su respaldo oficial y a la cantidad de años que hace que se definió, aún no se dispone de ninguna herramienta ampliamente difundida. Haberlas haylas, que en la última recopilación [8] aparecen hasta 26 productos; pero todos ellos son extremadamente complejos de instalación y de utilización. Por una parte requieren un profundo conocimiento de la norma X.500, por otra una gran cantidad de recursos (espacio en disco, capacidad de proceso y enlaces de comunicaciones).

Un servidor X.500 contiene información estructurada por campos (atributos), de forma que se pueden hacer consultas bastante elaboradas. La red X.500 se organiza jerárquicamente (de forma paralela a la jerarquía de dominios de mensajería) y permite a los clientes ir navegando para localizar lo deseado. X.500 permite navegar, pero no ayuda a navegar. Me explico: una cosa es poder llegar hasta la persona cuyos datos necesito (muy bien soportado por X.500) y otra saber dónde está. Lo habitual es tener unas coordenadas poco precisas y en esta circunstancia X.500 no prevé ninguna ayuda.

Hay desarrollos en curso que intentan superar estas dificultades. Por ejemplo, se está buscando un protocolo ligero [16] que prescinda de las partes más elaboradas del servicio X.500 y de una respuesta ágil a las preguntas más frecuentes de los clientes.

## **Whois**

Este es el servicio clásico en Internet [7]. Una serie de ordenadores a lo ancho del planeta ofrecen un servicio de directorio que abarca una cierta zona de

conocimiento. Originalmente había un sólo sitio que centralizaba toda la información sobre la Internet: nic.ddn.mil. Hace ya años que este lugar se vio desbordado y actualmente sólo proporciona información sobre recursos militares. El servidor básico de Internet es:

**telnet://rs.internic.net**, que proporciona información relativa a una cierta **clave**, que puede ser el nombre de algún responsable en Internet, un dominio, un número de red y poco más. Es básicamente un servidor de puntos de contacto.

Para personas en general, necesitamos consultar algún servidor más específico. Hay muchas instituciones, usualmente universidades, que proporcionan servidores whois. La relación oficiosa se mantiene en

`ftp://sipb.mit.edu/pub/whois/whois-servers.list`

A la mayor parte de los directorios se accede utilizando un cliente whois, por ejemplo:

`whois -h whois.dit.upm.es manas`

Desgraciadamente, cada servidor whois es un mundo aparte. Unos no saben de otros, y son incapaces de redirigir una consulta a un lugar más adecuado. Es más, cada servidor dispone la organización como mejor le parece: cada uno tiene sus propios campos con su propio significado y se requiere toda la inteligencia de un ser humano para interpretarlo correctamente. No es posible diseñar herramientas automáticas que analicen las respuestas de los servidores actuales.

Está en desarrollo un servicio denominado whois++ que establece un protocolo de comunicación entre servidores homogeneizados en cuanto a organización de contenidos. Hay una red jerárquica de servidores, de forma que cada uno proporciona a los demás una serie de pistas acerca de la información que poseen. Si un servidor recibe una pregunta y no sabe contestarla, sugiere servidores alternativos que pudieran hacerse cargo.

El servicio whois++ es un intento de superar tanto las limitaciones del whois (aislamiento de servidores) como la rigidez del X.500 (protocolo pesado y jerarquía estricta). Cabe imaginar un futuro donde los servidores sigan una norma X.500, se enlacen con el resto en base a un protocolo ligero, y sean capaces de interactuar con whois++ para enlazar inteligentemente los centros.

## **NetFind**

Este es un servicio de directorio basado en técnicas de inteligencia artificial que utiliza una serie de heurísticos y una extensa recopilación de datos dispersos por la red [15]. Dado un nombre y alguna pista sobre su lugar de trabajo, lanza una serie de consultas por la red localizando posibles respuestas.

Hay varios servidores netfind por la red. El original es accesible como

telnet://netfind@bruno.cs.colorado.edu

Hay varios otros servidores; si entra en dicho sistema, él mismo informa de los demás, por si quiere elegir otro más cercano.

Las sesiones con netfind son interactivas: él pide claves, aventura hipótesis, pide confirmaciones, consulta a otros servidores y, eventualmente, localiza el nombre de la cuenta y el ordenador en el que trabaja la persona que estamos buscando.

Aunque no es perfecta, se calcula que esta herramienta es capaz de localizar a unos 6 millones de personas. De hecho, es la mayor herramienta de que se dispone actualmente y la única capaz de combinar información proporcionada por diferentes servicios de páginas blancas.

## **Usenet**

Este es un directorio mantenido por el MIT que, durante años, ha venido recopilando direcciones de correo electrónico de gente que envía noticias al servicio de NEWS. El ámbito no es universal; pero permite localizar a los usuarios más activos de la red.

mail mail-server@pit-manager.mit.edu

El cuerpo del mensaje puede hacer solicitudes de búsqueda, por ejemplo:

send usenet-addresses/manas spain

## **Aspectos Prácticos**

Toda la teoría es inútil si no disponemos de formas prácticas de hacerla realidad. En esta sección intento proporcionar un breve catálogo de entradas maestras.

Debo empezar diciendo que la propia red contiene índices de sí misma, entre los que cabe destacar:

- <ftp://crl.dec.com/pub/misc/internet-index.txt>  
Una guía inspirada en el Harper's Index.
- [ftp://ftp.ripe.net/fyi/fyi\\_04.txt](ftp://ftp.ripe.net/fyi/fyi_04.txt)  
Intenta responder a las típicas preguntas de los nuevos usuarios de la red.
- [ftp://ftp.eff.org/pub/Net\\_info/Big\\_Dummy/](ftp://ftp.eff.org/pub/Net_info/Big_Dummy/)  
Un curso introductorio.

Existe por otra parte un libro excelente escrito por Ed Krol [10], que recoge un amplio catálogo de recursos disponibles.

Si se busca algo, una forma de empezar es vía el servicio de noticias (NEWS). Existen más de 10.000 grupos desde lo más general a lo más especializado. Si encuentra uno que responda a sus inquietudes, suscríbase. La relación íntegra de grupos puede obtenerse de

<ftp://ftp.uu.net/networking/news/config/newsroups>

Un grupo particularmente interesante es el denominado news.answers que distribuye las denominadas FAQs (Frequently Asked Questions), una especie de enciclopedia como las que se usaban antiguamente en los colegios. Las FAQs se renuevan periódicamente, y hay un número ingente de ellas, clasificadas por temas. Hay varios centros que las almacenan sistemáticamente. Por ejemplo:

<ftp://ftp.uu.net/usenet/news.answers/>  
o <ftp://rtfm.mit.edu/pub/usenet/news.answers/>

Hay una larga serie de ellas que cubren áreas tratadas en este artículo. Mis referencias se quedarán obsoletas en poco tiempo; pero estas FAQ le permitirán a cualquiera mantenerse al día. La mayoría no necesitan muchas explicaciones sobre su contenido; para las que no lo dejan claro, añado una breve descripción:

- active-newsgroups  
[\(Pulse aquí para traer el documento\)](#)
- communication-net-resources  
Es un catálogo de recursos accesibles por la red. Lo mantiene John December, un doctorando del RPI, y es muy amplio y actual.
- [\(Pulse aquí para traer el documento\)](#)
- finding-addresses  
Guía práctica para localizar direcciones de correo electrónico de personas accesibles en la red.
- [\(Pulse aquí para traer el documento\)](#)
- finding-groups  
[\(Pulse aquí para traer el documento\)](#)
- finding-sources  
Guía práctica para localizar programas (a veces, incluso fuentes) accesibles en la red.
- [\(Pulse aquí para traer el documento\)](#)
- fsp-faq  
FSP es una versión alternativa del clásico FTP. Es mucho más ágil, en particular sobre ordenadores poco potentes o sobre redes de baja capacidad; pero no está muy extendido.
- [\(Pulse aquí para traer el documento\)](#)
- ftp-list.faq  
[\(Pulse aquí para traer el documento\)](#)
- ftp-list.sitelist  
Relación de servidores anónimos de FTP, con una breve descripción de su contenido.
- [\(Pulse aquí para traer el documento\)](#)

- gopher-faq  
[\(Pulse aquí para traer el documento\)](#)
- internet-services.\*  
Describe servicios/protocolos y la manera de utilizarlos eficazmente.  
[\(Pulse aquí para traer el documento\)](#)
- mail.archive-servers\*  
Servidores a vuelta de correo. Se les manda un mensaje cuyo contenido son instrucciones a un robot en el centro receptor, robot que reacciona devolviendo la información que le pidamos.
- [\(Pulse aquí para traer el documento\)](#)
- mail.college-email  
Una descripción de cómo está organizado el espacio de direcciones de correo en muchas universidades, mayormente norteamericanas.
- [\(Pulse aquí para traer el documento\)](#)
- mail.country-codes  
Códigos ISO de identificación de países. Los del primer mundo nos los sabemos todos de memoria; pero también hay gente en el tercer mundo, y hay multitud de países recién creados en el este de Europa.
- [\(Pulse aquí para traer el documento\)](#)
- mail.inter-network-guide  
Guía práctica de cómo enviar un mensaje a personas accesibles en otras redes que no son parte integrante de Internet, y viceversa. Por ejemplo, para contactar con Compuserve, MCI-mail, ...
- [\(Pulse aquí para traer el documento\)](#)
- mail.mailing-lists  
Hay muchas listas de correo en las que se discuten temas o productos específicos. Un mensaje enviado a una lista se replica automáticamente para todos los que estén apuntados a ella.
- [\(Pulse aquí para traer el documento\)](#)
- msdos-archives.\*  
Programas para PCs.
- [\(Pulse aquí para traer el documento\)](#)
- newsgroups-reviews  
[\(Pulse aquí para traer el documento\)](#)
- periodic-postings  
[\(Pulse aquí para traer el documento\)](#)
- social-newsgroups  
[\(Pulse aquí para traer el documento\)](#)
- veronica.faq  
Veronica es una herramienta de localización de recursos (menús y ficheros) accesibles vía gopher.
- [\(Pulse aquí para traer el documento\)](#)
- wais-faq  
[\(Pulse aquí para traer el documento\)](#)
- www.faq  
[\(Pulse aquí para traer el documento\)](#)

Para localizar ficheros (documentos o programas, habitualmente) accesibles por medio de FTP se utiliza el servicio archie. Este es accesible en múltiples ordenadores, por ejemplo:

<telnet://archie@archie.doc.ic.ac.uk>  
<gopher://gopher.stanford.edu:4320/1archie>  
<http://hoohoo.ncsa.uiuc.edu/archie.html>  
<mailto:archie@archie.funet.fi>

Para localizar entradas en los menús de gopher se suele utilizar el servicio veronica. Este es muy similar a archie en el sentido de recorrer periódicamente los servidores gopher, recopilar sus menús, indexarlos y ofrecerlos para consultas. Sólo es accesible vía gopher y la forma más eficaz es entrar en <gopher://gopher.micro.umn.edu>, seguir la entrada Other Gopher and Information Servers y de aquí a Search Titles in Gopherspace using veronica.

Hay unas 500 bases de datos anunciadas públicamente y accesibles por medio de WAIS. Se puede recoger una relación íntegra en

<ftp://ftp.wais.com/pub/directory-of-servers/wais-sources.tar>  
<wais://quake.think.com/directory-of-servers>

El número de servidores gopher es tremendo y crece continuamente por lo fácil que es instalarlo. La gran mayoría están dados de alta en la Universidad de Minnesota, que es un buen sitio para empezar a buscar. Use gopher para ir a <gopher://gopher.micro.umn.edu>, siga la entrada Other Gopher and Information Servers y de aquí a All the Gophers in the World.

La explosión de servidores WWW es abrumadora y difícil de seguir al día. Al mismo tiempo, la posibilidad de poder entremezclar textos explicativos con enlaces activos ha hecho aparecer una pléthora de documentos que sirven de guía por la Internet. Creo que debo destacar

<ftp://ftp.rpi.edu/pub/communications/internet-cmc.html>

Fue escrito por un estudiante de doctorado y que es una guía exhaustiva de información sobre la Internet en sí, redes de ordenadores y temas relacionados. En particular le interesan los aspectos técnicos, sociales, cognitivos y sicológicos de las aplicaciones de la red.

<http://consultant-www.mit.edu/univ.html>  
Universidades.  
<http://www.ai.mit.edu/non.html>  
Organizaciones sin ánimo de lucro.  
<http://tns-www.lcs.mit.edu/commerce.html>  
Relación de empresas que ofrecen sus servicios vía WWW.

## Conclusiones

Internet es como una ciudad grande; pero no es una selva. Uno puede perderse si no utiliza guías adecuadas o si no es disciplinado, metódico y riguroso; pero no es peligroso. Hay gente mala, gente poco fiable, bromistas y estafadores; junto a investigadores, profesionales y comerciantes. Es simplemente el punto de encuentro de muchas culturas donde se puede encontrar casi de todo utilizando los medios adecuados, que son los expuestos en este artículo.

Muchas de las herramientas descritas son de dominio público. Esto las convierte en especialmente estables a medio y largo plazo, por su amplísima difusión, así como les da una oportunidad de llegar a estar casi libres de defectos pues se usan ampliamente y muchos contribuyen a depurarlas. Es posible que algunas vayan pasando a ser productos comerciales, añadiendo una buena documentación y un soporte profesional ante fallos. Es muy probable que en el futuro convivan servidores comerciales con clientes de dominio público: un mundo realmente abierto de interconexión de recursos.

## **Appendix A. Identificación Universal de Recursos**

La principal aportación del sistema WWW no es tanto la introducción del concepto de hipertexto, que ya existía en muchos sistemas, como su generalización para abarcar múltiples tipos de recursos, incluyendo:

1. La identificación del recurso
2. Su ubicación en la red: ordenador y directorio
3. La forma de acceso (el protocolo a usar)

Tim Berners-Lee propone la utilización de URIs (Universal Resource Identifiers) [1] para saber exactamente cómo se puede acceder a cierta información. URI es un concepto muy genérico y aún poco maduro; pero que ya se va concretando en una serie de normas de identificación eficaz en la Internet: los URLs (Universal Resource Locators). Para estos propone una notación concisa y legible, que se pretende universalizar. Es la que he venido utilizando a lo largo del texto.

Aún no están normalizados todos los mecanismos; pero los más habituales los recojo a continuación:

`ftp://host/path`  
`ftp://login:password@host:port/path`

Probablemente, el URL más utilizado en la actualidad. Identifica un ordenador, un directorio e incluso puede llegar a un fichero concreto. Es habitual entrar por el puerto 20, usando como *login* la palabra anonymous y como *password* la dirección de correo electrónico del que llama.

`http://host:port/path`

HTTP es el protocolo que comunica clientes y servidores WWW. Permite intercambiar documentos, negociando previamente las posibilidades de las

partes. Habitualmente trabaja en el puerto 80, bastando indicar el nombre del ordenador.

gopher://host:port/type selector

Gopher suele comunicarse a través del puerto 70, por lo que es habitual que se omita.

wais://host:port/database

wais://host:port/database?search

Los servicios de indexación permiten realizar consultas, habitualmente en el puerto 210.

telnet://login:password@host:port

Acceso vía terminal virtual.

mailto:mailbox@domain

Un mensaje electrónico.

news:group

Un grupo de news. Algunas aplicaciones ya entienden esta notación directamente, si bien la mayor parte de ellas requieren que el usuario desglose el URL y va transcribiendo la información correspondiente según la aplicación va pidiendo datos.

### **Bibliografía**

**Copyright © 1994-1998, ATI, Asociación de Técnicos de Informática.**

## **EVALUACION DE LOS PRINCIPALES "BUSCADORES" DESDE UN PUNTO DE VISTA DOCUMENTAL: RECOGIDA, ANÁLISIS Y RECUPERACION DE RECURSOS DE INFORMACION**

*Ángeles Maldonado Martínez. CINDOC-CSIC*

*Elena Fernández Sánchez. CINDOC-CSIC*

Resumen: El presente trabajo describe y analiza sistemáticamente las características documentales de los veinte buscadores, índices y motores de búsqueda, internacionales y en español de mayor relevancia. En primer lugar se explica el criterio aplicado para seleccionar los buscadores. A continuación se definen las funciones básicas que debe de realizar y las características que debe reunir un buscador como bases de datos documental que es desde el punto de vista de la recogida, análisis, búsqueda y recuperación de recursos. Seguidamente se describen pormenorizadamente cada uno de los buscadores seleccionados. En último lugar se analizan los resultados obtenidos.

Palabras-clave: Internet, buscadores, motores de búsqueda, índices, características documentales, recogida de información, análisis documental, recuperación de información.

### **1. OBJETIVO**

El objetivo del presente estudio es describir y analizar sistemáticamente las características documentales (recogida y análisis de información, búsqueda y resultados) de los veinte buscadores, índices y motores, internacionales y en español de mayor relevancia.

### **2. FASES DEL ESTUDIO**

Los pasos seguidos para la realización del presente trabajo han sido los siguientes:

- a. Definición de criterios para seleccionar los buscadores y selección de los mismos.
- b. Definición de las características documentales a cumplir por los buscadores seleccionados.
- c. Descripción de las características documentales, definidas con anterioridad, de cada uno de los buscadores seleccionados.
- d. Análisis de resultados a partir de los datos obtenidos.

### **1. SELECCIÓN DE BUSCADORES**

En la actualidad existe un gran número de buscadores, tanto índices temáticos como motores de búsqueda. Para realizar el presente estudio se han seleccionado veinte, diez de ámbito internacional y otros diez dirigidos al mundo hispano. Tanto en un caso como en otro los elegidos son aquellos que tienen un mayor número de URLs enlazadas. En un primer momento fueron preseleccionados aquellos buscadores más citados en diversos compiladores. Para realizar la selección definitiva se lanzó una búsqueda por "link" (localización de páginas enlazadas a una determinada dirección) en tres buscadores que permiten interrogar por este campo: Altavista.digital.com, Altavista.magallanes.net e Infoseek.

En el caso de los buscadores internacionales el ranking se estableció en función de los resultados obtenidos en Altavista.digital.com. Los buscadores seleccionados y el número de páginas enlazadas a cada uno de ellos fueron los siguientes: Yahoo! (754.739), Altavista (272.761), Excite (206.979), Lycos (180.128), WebCrawler (93.904), Hotbot (93488), Magellan (60.994), Infoseek (33.214), Galaxy (26.528) y Looksmart (10.433).

Con respecto a los buscadores en español (ver gráfico 2) se primaron los datos obtenidos en Altavista.magallanes.net y los resultados fueron los siguientes: Ole (5.770), Donde? (3.673), Ozu.com (2.072), Elcano (1742), Biwe (1.415), Ozu.es (1.229), Trovator (1.155), Tarántula (1.056), Hispavista (869) y Sol (706).

## 2. DEFINICIÓN DE CARACTERÍSTICAS DOCUMENTALES

Los buscadores son las herramientas más importantes para localizar información en Internet. Tanto los índices temáticos como los motores de búsqueda son bases de datos de URLs. En el caso de este trabajo se han establecido dos tipos de URLs: las que representan páginas individualizadas y las que representan recursos de información completos compuestos por una o más páginas. Por norma general los índices gestionan recursos de información y los motores páginas

Las bases de datos documentales deben realizar unas funciones básicas y cumplir una serie de características, que a los buscadores también se les debe exigir. A continuación se describen dichas funciones y caracteristicazas agrupadas en tres apartados: recogida y análisis de información, búsqueda y resultados.

### a. RECOGIDA Y ANALISIS

En una cualquier base de datos la captación de datos se puede hacer de forma manual o de forma automática. Lo mismo ocurre en el caso de los buscadores. El procedimiento manual, utilizado por los índices, consiste en llenar un formulario de alta que los propios buscadores proporcionan. Los datos solicitados en estos cuestionarios varían de unos a otros. Cuanto más

completa y pormenorizada sea la toma de datos, las posibilidades de búsqueda podrán ser también más completas. En este trabajo se han considerado dos tipos de datos a consignar, unos que se les ha otorgado la categoría de imprescindibles: título, URL, clasificación y descripción; y otros considerados recomendables: palabras-clave, localización, idioma, tipo de información (información académica, información comercial, directorio, publicación electrónica, etc.) y tipo de propietario (organismos públicos, empresas, páginas personales, etc.). En el caso de los recursos de información existentes en Internet parece excesivo utilizar las cuatro modalidades de representación de contenido: título, clasificación, palabras-clave y descripción (resumen). La clasificación viene avalada por la utilidad de las búsquedas a través del índice temático. Entre el título, la descripción y las palabras clave, se han considerado como imprescindibles las dos primeras: el título por ser la modalidad más concisa y por lo tanto la que introduce un grado de error menor en la recuperación y la descripción, porque en la totalidad de los casos su longitud está muy limitada, pudiéndose ser considerada como un conjunto de palabras clave unidas sintácticamente. Además el contenido de la descripción se puede utilizar posteriormente en la presentación de resultados. Sin embargo también se ha considerado recomendable la existencia de palabras clave, por aumentar la pertinencia en la recuperación.

El procedimiento de toma de datos automático es el que normalmente utilizan los motores de búsqueda, quienes únicamente solicitan la URL de la página a dar de alta. A continuación su robot es capaz de visitar e indizar la página dada de alta y todas aquellas otras que están enlazadas. El lenguaje HTML permite utilizar una serie de etiquetas, denominadas Meta, a través de las cuales se puede añadir cualquier tipo de información sobre una página, en el momento de su confección. Con ayuda de estas etiquetas se puede realizar un análisis completo de una URL desde el punto de vista documental. Principalmente se suelen utilizar para describir el contenido a través de pequeños resúmenes y palabras-clave. Hay robots que son capaces de identificar las etiquetas META y extraer la información de las mismas para ser usada en la búsqueda o en la presentación de resultados. Sin embargo, existen otros que no la detectan o no utilizan su contenido. Los datos imprescindibles y recomendables a consignar en las etiquetas Meta serían los mismos que en el caso de los formularios de alta de los índices: título, URL, clasificación, descripción, palabras-clave, localización, idioma, tipo de recurso y propietario de las páginas. Hay que resaltar que las etiquetas Meta no son obligatorias a la hora de diseñar una página, y en la actualidad hay muchas URLs que carecen de ellas.

En documentación, en términos generales, son más aconsejables los sistemas que son capaces de analizar más pormenorizadamente las características formales y de contenido de cada uno de los registros que gestionan. En el caso de los buscadores, no se deben contraponer de

forma general índices y motores de búsqueda, sino más bien hay que indagar las características particulares tanto de unos como de otros.

Consecuentemente con lo anterior, si hubiera que establecer un ranking de categorías de buscadores desde el punto de vista de sistema de recogida y análisis de direcciones el orden sería el siguiente: 1) Índice temático con al menos título, URL, clasificación, descripción de cada una de las páginas. 2) Buscador con reconocimiento y análisis de etiquetas Meta. 3) Buscador sin reconocimiento y uso de las etiquetas Meta.

*b. BUSQUEDA*

Un sistema de búsqueda es mejor cuanto más flexible es y cuantas más posibilidades de recuperación ofrece. Las características que se han considerado básicas desde el punto de vista de la recuperación de la información que debe cumplir un buscador, ya sea un índice o un motor de búsqueda, son las que a continuación se citan:

- Formularios de búsqueda. Posibilidad de elegir entre un formulario simple y otro más detallado. El simple da la posibilidad de acercarse a los buscadores a aquellas personas que no están muy familiarizadas con el uso de las bases de datos. Los formularios más completos dan la oportunidad de realizar búsquedas más complejas a los usuarios expertos.
- Herramientas de búsqueda. Posibilidad de uso del truncado de palabras, de los operadores booleanos: AND, OR y el NOT y del paréntesis. También es recomendable la localización de términos compuestos (utilizando comillas o buscando frases). El contar con todas estas herramientas permite realizar búsquedas que definen muy ajustadamente la información que se requiere.
- Clasificación temática. Existencia de un índice general, pues facilita mucho la localización de información a aquellos que no saben concretar su tema de búsqueda y prefieren explorar los apartados temáticos propuestos por el sistema.
- Campos de búsqueda. La posibilidad de dirigir las búsquedas a campos determinados aumenta notablemente la pertinencia en la recuperación de información. Las búsquedas en texto libre son más exhaustivas pero menos pertinentes. La búsqueda en campos determinados está muy relacionada con la segmentación de datos en la recogida de datos. Cuanto más sectorializada esté la información extraída de un recurso de información y más se pueda concentrar la búsqueda a un tipo de dato concreto más pertinente será la recuperación. Como en el caso de la toma de datos se consideran campos imprescindibles: título, URL, descripción; y campos recomendables: palabras-clave, localización, idioma, tipo de información y tipo de propietario.

- Control del vocabulario. Uno de los problemas fundamentales en la recuperación de información es el control de vocabulario. Eliminar las sinonimias y las polisemias es una labor a realizar en todo sistema documental que se precie, pues disminuye considerablemente el ruido y el silencio.
- Detección de novedades. En todo sistema documental es imprescindible poder detectar los nuevos registros incorporados. Existen varios sistemas a través de los que detectar las novedades: creación de ficheros aparte, uso de etiquetas especiales, acotar por fecha de alta y ordenar los recursos por fecha de alta.

## C) RESULTADOS

Es aconsejable que los usuarios puedan elegir entre diferentes formatos de presentación de resultados o incluso diseñarlos a la medida. También es recomendable que todos los documentos que componen un sistema documental vayan acompañados de una descripción de su contenido. En cuanto al orden de presentación de los resultados, es importante que se pueda seleccionar entre varios criterios de ordenación. En el caso de los motores de búsqueda, ya que gestionan URLs, es recomendable que puedan presentar las páginas recuperadas agrupadas por servidores.

### 1. DESCRIPCION Y ANALISIS DE LOS BUSCADORES SELECCIONADOS

A continuación se describen las funciones y las características documentales de cada uno de los buscadores seleccionados. También se resalta lo positivo y lo negativo de cada uno de ellos.

#### YAHOO! (Yahoo! Inc.)

<http://www.yahoo.com/>

**Contenido:** cobertura internacional con nodos locales que sirven a sus respectivos ámbitos geográficos en: Alemania, Asia, Australia, Canadá, Corea, China, Dinamarca, España, Estados Unidos, Francia, Italia, Japón, Noruega, Reino Unido y Suecia.

**Recogida de Páginas:** la mayoría de los 'sites' son sugeridos por los propios usuarios, y más tarde son visitados y evaluados por un equipo de Yahoo!. Existe un formulario de sugerencias donde se solicita la URL, el título y la descripción del recurso a incluir. Yahoo! está organizado en categorías temáticas, por lo que el solicitante además debe indicar la(s) categoría(s) en las que piensa debe enmarcarse el recurso sugerido.

**Búsqueda:** se puede buscar siguiendo dos procedimientos: o bien a través de la clasificación temática o bien utilizando los formularios de búsqueda. En el primer caso se puede descender en la jerarquía hasta donde se considere oportuno y visualizar los recursos de interés. En el segundo, se teclean los términos en una ventana debiéndose elegir entre visualizar los epígrafes clasificatorios o los recursos que contengan dichos términos. Si el resultado no es satisfactorio, se tiene la posibilidad de lanzar la pregunta en las páginas indizadas por el motor de búsqueda Inktomi o conectar con otro buscador (Altavista, Hotbot, Infoseek, etc.). El sistema admite truncar, buscar por términos compuestos y el uso de los operadores booleanos AND (+), NOT (-) y OR. La opción 'Search Options' permite hacer una acotación en función del tiempo transcurrido desde el alta del recurso, preguntar por una URL determinada o buscar los términos en el título del recurso.

**Resultados:** no se puede formatear el resultado de una búsqueda. De cada 'site' se visualiza el título con el enlace, una pequeña descripción y el epígrafe clasificatorio al que pertenece, en el caso de que la búsqueda se haya realizado en la base de Yahoo!. Si la búsqueda se ha lanzado en Inktomi la descripción se corresponde con las primeras líneas de las páginas. Alcanzarán un mayor rango en la presentación aquellos documentos que contengan más palabras clave, que estas estén en determinadas partes del documento (se prima el título) o que el recurso pertenezca a una categoría que esté situada en una posición elevada, es decir más general.

**Ayudas:** las ayudas detallan la forma de realizar las búsquedas con ejemplos, los operadores a utilizar y los criterios de relevancia que se aplican en la presentación de resultados.

**Otros:** a las nuevas incorporaciones de les añade la etiqueta de 'nuevo' y se las posiciona al principio de cada epígrafe durante la primera semana.

#### **Positivo:**

- 1) El alta de recursos se realiza a través de un formulario donde se solicita la URL, el título y la descripción del mismo.
- 2) Existe una clasificación temática.
- 3) Posibilidades de búsqueda satisfactorias ya que se pueden utilizar los operadores AND, OR y NOT, se puede truncar y localizar términos compuestos.
- 4) Se pueden acotar las búsquedas a la URL y al título. Las búsquedas en texto libre se realizan sobre la descripción del recurso.
- 5) Se recuperan unidades de información y no páginas.
- 6) las ayudas son satisfactorias.
- 7) Se pueden detectar las nuevas incorporaciones al acotar por fecha de alta del recurso.
- 8) Tiene versión en español.

**Negativo:**

- 1) Posibilidades de búsqueda medianamente satisfactorias, ya que aunque el sistema permite truncar, localizar términos compuestos, y utilizar los operadores AND, OR y NOT, no es posible utilizar paréntesis.
- 2) No se controla el vocabulario.
- 3) No se pueden diseñar los formatos de presentación.
- 4) No se pueden mantener abiertas paralelamente la lista de resultados de la búsqueda y una de las direcciones recuperadas.

**ALTAVISTA** (Digital Equipment Corporation)

<http://altavista.digital.com/>

**Contenido:** cobertura internacional con nodos locales que sirven a sus respectivos ámbitos geográficos y lingüísticos en los siguientes países: Australia, Canadá, España, Estados Unidos, Japón y Suecia.

**Recogida de Páginas:** las páginas son recogidas por un robot que es capaz de indizarlas y de saltar a aquellas otras enlazadas con ellas. Las altas se realizan a partir de un formulario donde únicamente se solicita la dirección de la página a dar de alta. El sistema es capaz de reconocer las etiquetas META de título, descripción y palabras-clave y extraer la información que contienen.

**Búsqueda:** presenta dos formularios de búsqueda con casi las mismas prestaciones, ya que se puede acotar por idioma, buscar en campos concretos (dominios, servidores, URL's, links, imágenes, título e idioma), utilizar operadores booleanos (AND, OR, NOT), de proximidad y paréntesis, localizar términos compuestos y truncar. Los dos también contemplan la posibilidad de 'Afinar' la búsqueda, es decir hacer una segunda selección entre las páginas recuperadas originalmente en función de la elección de nuevos términos propuestos por el sistema por el grado de coincidencia con los demandados en un principio. La diferencia entre los dos formularios reside en que en el simple las páginas seleccionadas son presentadas por un orden de relevancia establecido por el propio sistema y en el avanzado es el usuario el que decide el orden de presentación haciendo uso de la opción 'Ranking', indicando el término que deben tener las páginas presentadas en primer lugar. En el formulario para búsquedas avanzadas se puede seleccionar documentos que hayan sido publicados o modificados por última vez en un periodo de tiempo establecido.

**Resultados:** no se puede formatear el resultado de una búsqueda. De cada página se visualiza: el título con el enlace, la URL y las primeras palabras en el caso de que no contenga etiqueta META de descripción, sí es así aparece el contenido de la misma. Además, aparece el tamaño de la página

(bits), el idioma y la fecha de la última modificación. Los criterios de relevancia para la ordenación de las páginas de una búsqueda simple son: si las palabras o frases demandadas se encuentran en las primeras líneas de una página; la frecuencia de ocurrencia; si las palabras o frases se una consulta múltiple se encuentran cerca una de otra en el documento. Se puede mantener abiertas simultáneamente la página de resultados y uno de los documentos presentados.

**Ayudas:** las ayudas detallan con ejemplos la forma de realizar las búsquedas, los operadores a utilizar y los criterios de relevancia que se aplican en la presentación de resultados.

**Otros:** Altavista conecta desde su página principal con 'Looksmart', índice de páginas clasificados en categorías.

**Positivo:**

- 1) Reconoce las etiquetas META y utiliza su información para indizar las páginas y como criterio de relevancia a la hora de presentación de resultados.
- 2) Las posibilidades de búsqueda son bastante completas: truncado, términos completos, operadores booleanos (AND, OR, NOT) y paréntesis.
- 3) Posibilidad de buscar por campos concretos: dominios, servidores, URL's, links, imágenes, título e idioma.
- 4) Posibilidad de "afinar" la búsqueda utilizando términos propuestos por el sistema.
- 5) Posibilidad de mantener una página abierta a la vez que se visualiza el listado de resultados.
- 6) Las ayudas son bastante completas.
- 7) Se puede recuperar páginas publicadas o revisadas en un periodo de tiempo dado.
- 8) Versión en español.

**Negativo:**

- 1) Carece de clasificación temática.
- 2) No se puede recuperar información utilizando los campos de descripción y palabras clave.
- 3) No se controla el vocabulario.
- 4) Se recuperan páginas y no recursos de información que además no pueden ser agrupadas por servidores.
- 5) No se pueden diseñar los formatos de presentación.

**EXCITE (Excite Inc.)**

<http://www.excite.com/>

**Contenido:** cobertura internacional con versiones locales en Alemania, Francia, Holanda, Suecia, Reino Unido, Japón, China y Australia.

**Recogida de páginas:** las páginas que contiene su base de datos son localizadas e indizadas por un robot, aunque también se pueden dar de alta usando un formulario existente para tal fin. El robot desestima toda la información proveniente de las etiquetas META.

**Búsqueda:** existen dos formularios de búsqueda: uno con una única ventana y otro más completo. Las búsquedas pueden realizarse a través de palabras o de términos compuestos que pueden ser combinados entre sí con operadores booleanos (AND, OR y NOT). Se pueden utilizar paréntesis y truncar. No se puede limitar por campos. En la búsqueda aparecen dos funciones interesantes, por una parte la sugerencia de palabras asociadas a los términos solicitados (Search Wizard) y por otra la posibilidad de pedir documentos con las mismas características de aquel o aquellos que más se ajustan a las preferencias del solicitante (More Documents like this one). En la presentación y ayudas se indica que Excite realiza las búsquedas por conceptos, es decir que ante una búsqueda, el sistema identifica los conceptos representados por las palabras utilizadas y es capaz de localizar esos mismos conceptos. Lo que no explica las ayudas es el procedimiento exacto para realizar esta búsqueda inteligente.

**Resultados:** los resultados se presentan según un orden de relevancia establecido por el sistema. De cada una de las páginas seleccionadas se puede elegir entre visualizar únicamente el título y el enlace o además la descripción (primeras palabras de la página), la URL y el % de relevancia. Las páginas pueden ser agrupadas por servidores (List by Web Site).

**Ayudas:** en las ayudas se detalla convenientemente como realizar una búsqueda y los criterios de presentación de recursos.

**Otros:** también presenta una selección de páginas clasificadas por temas

**Positivo:**

- 1) Se pueden utilizar los operadores AND, OR y NOT, utilizar los paréntesis, truncar y localizar términos compuestos.
- 2) Posibilidad de 'afinar' la búsqueda utilizando términos asociados propuestos por el propio sistema.
- 3) Se pueden pedir documentos con las mismas características a uno anteriormente seleccionado.
- 4) Realiza un cierto control del vocabulario en los términos de búsqueda.
- 5) A la hora de presentación de resultados se puede elegir entre un formato reducido (sólo título) y uno con mayor información.
- 6) Las páginas recuperadas pueden ser agrupadas por servidores.

7) Las ayudas proporcionan suficiente información.

**Negativo:**

- 1) Las páginas clasificadas temáticamente representan una pequeña parte del total.
- 2) No reconoce las etiquetas META.
- 3) No se puede buscar por campos, salvo por idioma.
- 4) La descripción de las páginas se corresponde con las primeras palabras de la página.
- 5) Se recuperan páginas y no recursos, aunque estas puedan ser agrupadas por servidor.
- 6) No se pueden mantener varias páginas abiertas a la vez que se visualiza la lista de resultados.
- 7) No hay forma de detectar las páginas nuevas.
- 8) No existe versión en español.

**LYCOS** (Lycos Inc. Carnegie Mellon University)

<http://www.lycos.com/>

**Contenido:** cobertura internacional con nodos locales en: Alemania, Bélgica, España, Estados Unidos, Francia, Holanda, Italia, Reino Unido, Suecia y Suiza.

**Recogida de Páginas:** las páginas son localizadas por un robot a partir de las URL's dadas de alta en el formulario que existe para tal fin, donde sólo se solicita la URL y el e-mail de la persona que realiza el alta. Ignora las etiquetas META de descripción y palabras clave.

**Búsqueda:** cuenta con un formulario simple y otro más complejo (Lycos Pro). En ambos se puede acotar por idioma y utilizar: operadores booleanos (AND, OR y NOT), operadores de proximidad y paréntesis. Sin embargo, no se admite el truncado. Además en el formulario complejo se puede elegir entre buscar los términos de búsqueda en todo el documento, el título o la URL. También se pueden hacer búsquedas en un web determinado existente en Lycos, indicando su dominio en el formulario de búsqueda. La opción "más resultados" permite visualizar las páginas de iguales características que una seleccionada. Se pueden localizar imágenes y sonidos.

**Resultados:** se puede elegir entre visualizar las páginas seleccionadas agrupadas por dominios o hacerlo individualizadamente. De cada página se visualiza: el título con el enlace, la URL, las primeras palabras a modo de resumen, el tamaño en bits de la página y el grado de pertinencia. El usuario puede elegir entre varios el criterio de relevancia a utilizar para el orden de presentación de las páginas.

**Ayudas:** las ayudas detallan la forma de realizar las búsquedas, los operadores a utilizar y los criterios de relevancia a aplicar en la presentación de resultados.

**Otros:** además de la base de datos confeccionada a partir del motor de búsqueda, Lycos contiene lo que denomina Webguide, que no es más que un índice de recursos seleccionados clasificados por temas. No tiene motor interno de búsqueda.

**Positivo:**

- 1) Se puede buscar por los siguientes campos: idioma, título y URL.
- 2) Se pueden hacer búsquedas dentro de un web determinado existente en Lycos.
- 3) Visualización de páginas con igual características que una seleccionada.
- 4) Se pueden localizar imágenes y sonidos.
- 5) Se recuperan páginas, pero pueden ser agrupadas por dominios.
- 6) El usuario puede elegir los criterios de relevancia en el orden de presentación de los resultados.
- 7) Las ayudas son bastante completas.
- 8) Existe versión en español.

**Negativo:**

- 1) No reconoce las etiquetas META de palabras claves ni de descripción, por lo tanto su contenido no puede ser utilizado ni en la indización ni en la presentación de resultados.
- 2) Sólo están clasificadas partes de las páginas.
- 3) Las posibilidades de búsqueda son incompletas, pues no admite el truncado.
- 4) No se controla el vocabulario.
- 5) En los resultados la descripción de la página se realiza a través de las primeras líneas del texto de la misma.
- 6) No se pueden detectar las novedades.

**WEBCRAWLER (Excite Inc.)**

<http://www.webcrawler.com/>

**Contenido:** cobertura internacional.

**Recogida de páginas:** los usuarios pueden sugerir nuevas páginas a partir de un formulario, advirtiéndose que de un web únicamente se consigne las URLs de la página principal y aquellas otras de las secciones relevantes. Webcrawler indiza todas las palabras de una página hasta 1 MB de texto.

No utiliza la información de las etiquetas META. Expertos de Webcrawler confeccionan un directorio ("The WebCrawler Guide") donde una selección de "sites" se encuentra organizada en categorías.

**Búsqueda:** las búsquedas se realizan a través de un formulario con una única ventana donde deben consignarse los términos a localizar. Las búsquedas pueden realizarse a través de unitérminos o de términos compuestos que pueden ser combinados entre sí con operadores booleanos (AND, OR, NOT). Se pueden utilizar paréntesis. No se puede limitar por campos. Al igual que en Excite y Magellan se pueden pedir documentos con las mismas características de aquel o aquellos que más se ajustan a las preferencias del solicitante ("Similar Pages"). También se puede consultar el directorio y seleccionar páginas descendiendo en la jerarquía hasta encontrar el epígrafe adecuado.

**Resultados:** existen dos formatos de presentación: uno exclusivamente con los títulos de las páginas y otro, además de con el título, con la URL, las primeras palabras a modo de descripción y el grado de relevancia. En ambos casos los resultados se ordenan según su grado de relevancia estimado a partir del número de palabras solicitadas encontradas en el documento. En el caso del directorio, en primer lugar aparecen las páginas recomendadas por WebCrawler y a continuación el resto ordenadas alfabéticamente con una breve descripción del contenido de las mismas.

**Ayudas:** en las ayudas se detalla como funciona el sistema y como realizar una búsqueda convenientemente. Se presentan ejemplos.

#### **Positivo:**

- 1) Las posibilidades de búsqueda son totalmente satisfactorias, pues se pueden utilizar los operadores AND, OR y NOT y los paréntesis, se puede truncar y localizar términos compuestos.
- 2) se pueden localizar páginas con las mismas características que una predeterminada.
- 3) Se puede elegir entre dos formatos de presentación.
- 4) los resultados se ordenan por criterios de relevancia.
- 5) Las ayudas son bastante completas.

#### **Negativo:**

- 1) No utiliza la información de las etiquetas META.
- 2) Las páginas están parcialmente clasificadas.
- 3) No se puede limitar por campos.
- 4) Se recuperan páginas que no pueden ser agrupadas por servidores.
- 5) La descripción de la página viene dada por las primeras líneas de la misma.
- 6) No se pueden detectar las novedades.

7) No existe versión en español.

**HOTBOT** (Wired Digital Inc.)

<http://www.hotbot.com/>

**Contenido:** cobertura internacional.

**Recogida de Páginas:** las páginas son localizadas por un robot a partir de las URL's dadas de alta en un formulario, donde debe indicarse la URL y el e-mail de la persona que la realiza. El robot extrae información de las etiquetas META de título, descripción y palabras clave.

**Búsqueda:** cuenta con dos formularios de persiana: uno simple y otro más complejo. En el primero se puede acotar por continente y fecha de alta. En el más complejo, además de por los conceptos anteriores, se añaden más elementos. En ambos casos se pueden utilizar las siguientes Meta words (etiquetas de campo): dominio, título, links, fechas de creación y modificación de las páginas. También se pueden localizar páginas dentro del servidor designado. Reconoce términos compuestos y se pueden utilizar: operadores booleanos (AND, OR y NOT), operadores de proximidad, paréntesis y truncado.

**Resultados:** se puede elegir entre tres formatos de visualización. En el formato completo de cada página se visualiza: el título con el enlace, la URL y las primeras palabras a modo de resumen o en el caso de contar con etiqueta META de descripción del contenido existente en la misma. Además aparece el tamaño de la página (bits), la fecha de alta en el sistema y el grado de relevancia. Las páginas se presentan por orden de relevancia donde se tiene en cuenta la posición de los términos solicitados (título y etiqueta Meta de palabras-clave), la frecuencia en la página y la profundidad de la misma. Hotbot es capaz de mantener simultáneamente la conexión con el listado de resultados y con una de las páginas recuperadas.

**Ayudas:** las ayudas detallan con ejemplos la forma de realizar las búsquedas, los operadores a utilizar y los criterios de relevancia que se aplican en la presentación de resultados.

**Otros:** Hotbot enlaza con el índice de páginas clasificados "Looksmart" a partir de la presentación de sus categorías.

**Positivo:**

1. Reconoce las etiquetas META de título, descripción y palabras clave.

2. Las posibilidades de búsqueda son muy completas, ya que se pueden utilizar todos los operadores booleanos y los paréntesis, se puede truncar y localizar palabras compuestas.
3. Se puede acotar por dominio, título, continente, país, fecha de alta o modificación de las páginas, link y por numerosos campos que tienen que ver con las características informáticas de las páginas.
4. Se puede elegir entre tres formatos de visualización.
5. Para la descripción de las páginas se utiliza el contenido de esta etiqueta Meta en el caso de que exista.
6. El orden de presentación de resultados obedece a unos criterios de relevancia.
7. Se puede mantener conexión simultáneamente con el listado de resultados y con una de las páginas recuperadas.
8. Las novedades pueden ser detectadas utilizando el campo de fecha de alta y modificación.
9. Las ayudas son bastante completas.

**Negativo:**

- 1) Las páginas no están organizadas por clasificación temática.
- 2) No se puede acotar ni por descripción, ni por palabra clave.
- 3) No se controla el vocabulario.
- 4) Se recuperan palabras y no recursos de información y las páginas no pueden ser agrupadas por recursos.
- 5) No existe versión en español.

**MAGELLAN** (The McKinley Group, Inc., a subsidiary of Excite Inc)

<http://www.mckinley.com/>

**Contenido:** cobertura internacional.

**Recogida de páginas:** los usuarios pueden sugerir nuevas páginas a partir de un formulario en el que únicamente deben consignar la URL de la página a incluir. Las páginas son visitadas y dadas de alta definitivamente por un equipo de expertos de Magellan.

**Búsqueda:** Magellan está organizado en categorías en las que se encuentran enmarcadas las páginas visitadas y evaluadas por los expertos. Existen dos procedimientos para realizar las búsquedas: o bien descendiendo por las categorías de su clasificación hasta llegar al epígrafe adecuado o bien tecleando los términos deseados sobre un formulario de búsqueda. En este formulario se debe especificar el conjunto de documento sobre el que se quiere buscar: documentos revisados por Magellan, documentos con luz verde o en la base de datos de Excite (The Entire Web). No funciona el truncado. Las búsquedas pueden realizarse a través

de unitérminos o de términos compuestos que pueden ser combinados entre sí con operadores booleanos (AND, OR y NOT). Se pueden utilizar paréntesis. No se puede limitar por campos. Al igual que en Excite se pueden pedir documentos con las mismas características de aquel o aquellos que más se ajustan a las preferencias del solicitante (Find Similar). También como en Excite en la presentación se indica que el sistema realiza las búsquedas por conceptos, es decir que ante una pregunta el sistema identifica los conceptos representados por las palabras utilizadas y es capaz de localizar estos mismos conceptos en otras páginas. Lo que no se explica es el procedimiento exacto para realizar esta búsqueda inteligente.

**Resultados:** los resultados se presentan según un orden de relevancia establecido por el sistema, que no se explica en las ayudas. De cada una de las páginas seleccionadas aparece el título con el enlace, un pequeño resumen que sirve como descripción de la página, la URL y el % de relevancia.

**Ayudas:** en las ayudas se detalla convenientemente como realizar una búsqueda presentándose ejemplos.

**Positivo:**

1. La selección de páginas la hace un equipo de expertos del propio Magellan.
2. Todos los recursos de información existentes en la base están clasificados temáticamente.
3. Existe cierto control del vocabulario.
4. Se pueden pedir documentos de iguales características a uno designado.
5. Se recuperan unidades de información y no páginas.
6. La descripción del recurso se hace a través de un pequeño resumen.
7. Los recursos se ordenan utilizando criterios de relevancia.
8. Las ayudas a la búsqueda son completas, pero no se explican ni los criterios de indización ni los de relevancia para establecer el orden de presentación de recursos.

**Negativo:**

- 1) Las posibilidades de búsqueda no son completas, pues no se puede truncar.
- 2) No se puede buscar por campos concretos.
- 3) No se pueden seleccionar formatos.
- 4) No hay forma de detectar las novedades.
- 5) No existe versión en español.

<http://guide.infoseek.com/>

**Contenido:** cobertura internacional con versiones locales en España, Alemania, Francia, Italia, Dinamarca, Suecia, Reino Unido y Japón.

**Recogida de Páginas:** las páginas que contiene su base de datos son localizadas e indizadas por un robot, también se pueden dar de alta usando un cuestionario que el propio sistema presenta donde exclusivamente se solicita la URL de la página. Las páginas se dan de alta individualizadamente. El robot es capaz de identificar las etiquetas META de palabras clave ("keywords") y de descripción existentes en las páginas utilizando la primera para la indización de la página y la segunda como resumen del contenido en la visualización de resultados.

**Búsqueda:** la búsqueda se puede realizar utilizando cualquiera de los dos formularios existentes: uno para búsquedas simples y otro para las más complejas. El sistema permite localizar palabras y frases que pueden ser combinadas con los operadores AND, OR y NOT, pero no se puede utilizar paréntesis. Se pueden truncar términos. Las búsquedas pueden hacerse en texto libre o bien limitarse a los siguientes campos: URL, "hyperlinks", título y lugar geográfico.

**Resultados:** los resultados son presentados siguiendo un orden de relevancia primándose aquellas páginas en las que los términos requeridos aparecen en el título o cerca del inicio, contienen un mayor número de ocurrencias o los términos solicitados son considerados relevantes (palabras poco comunes en una base de datos). De cada página seleccionada aparece el título con el enlace, la URL, la descripción (la que aparece en la etiqueta META "descripción", en el caso de que exista, o las primeras líneas que aparecen en la página propiamente dicha), el tamaño de la página (bits) y la fecha de alta en el sistema.

**Ayudas:** en las ayudas se detalla convenientemente como realizar una búsqueda y los criterios de presentación de recursos.

**Otros:** también presenta una sección de páginas clasificadas directamente por un equipo de personas de Infoseek.

**Positivo:**

1. Reconoce las etiquetas META de palabras-clave y de descripción.
2. Posibilidades de búsqueda bastante completas: truncado, búsqueda de términos compuestos y operadores booleanos (AND, OR, NOT).
3. Posibilidad de buscar por los siguientes campos: URL, "hiperlink", título y lugar geográfico.

4. Las páginas recuperadas se presentan siguiendo un orden de relevancia.
5. Se utiliza el texto de la etiqueta META de descripción para resumir el contenido de cada página recuperada.
6. Detección de novedades por la fecha de alta en el sistema.
7. Las ayudas son bastante completas.
8. Versión en español.

**Negativo:**

- 1) El alta debe hacerse página a página.
- 2) Las páginas se encuentran parcialmente clasificadas.
- 3) No permite el uso del paréntesis.
- 4) No se puede buscar ni en el campo de palabras-clave ni en el de descripción.
- 5) No se controla el vocabulario.
- 6) Existe un único formato de presentación.
- 7) Se recuperan páginas y no recursos que además no pueden ser agrupadas por servidores.
- 8) Se conoce la fecha de alta de las páginas, pero no se puede acotar por este criterio.

**GALAXY** (CyberGuard Corp.)

<http://www.einet.net>

**Contenido:** cobertura internacional.

**Recogida de Páginas:** las altas se producen a partir de un formulario donde se solicita título, URL, idioma, tipo de información (institución académica, organización comercial, directorio, revista, etc.), localización geográfica y descripción.

**Búsqueda:** existen dos procedimientos de búsqueda: utilizar uno de los dos formularios que existen o descender por la clasificación temática. El formulario simple tan sólo es una ventana donde se teclean los términos de búsqueda. El formulario destinado a las búsquedas avanzadas está compuesto de tres ventanas: una para los términos, otra para restringir a un epígrafe de la clasificación y un tercero para elegir el formato de salida. También se puede elegir entre buscar exclusivamente en el título y la URL o hacerlo en el registro completo. Se puede truncar y localizar términos compuestos y utilizar los operadores lógicos AND, OR y NOT.

**Resultados:** se puede elegir el formato de visualización entre tres opciones. El formato más completo contiene la siguiente información: título con el enlace, grado de pertinencia, descripción, palabras más frecuente dentro del recurso, tipo de recurso, clasificación, URL, fecha de alta en el

sistema y fecha de indización. El orden de presentación obedece a ciertos criterios de relevancia que no son explicados.

**Ayudas:** se explica convenientemente como realizar las búsquedas y como llenar el formulario de alta de nuevos recursos, pero no se da información sobre como funciona el sistema ni cuáles son los criterios de relevancia en el orden de presentación de recursos.

**Positivo:**

1. Las altas se realizan a partir de un formulario donde se solicitan: título, URL, idioma, tipo de información (institución académica, organización comercial, directorio, revista, etc.), localización geográfica y descripción.
2. Los recursos están todos clasificados temáticamente.
3. Las posibilidades de búsqueda son casi completas pues se pueden utilizar todos los operadores lógicos (AND, OR y NOT), se puede truncar y buscar términos compuestos. No se pueden utilizar paréntesis.
4. Se recuperan unidades de información y no páginas.
5. Se puede seleccionar entre tres formatos de presentación.
6. La información aportada para cada registro recuperado es muy amplia, pues en el formato completo se reproducen los siguientes datos: título con el enlace, grado de pertinencia, descripción, palabras más frecuente dentro del recurso, tipo de recurso, clasificación, URL, fecha de alta en el sistema y fecha de indización.
7. Los recursos recuperados se ordenan en función de ciertos criterios de relevancia.
8. Se pueden detectar las novedades, pues existe un fichero especial donde se almacenan las nuevas incorporaciones.

**Negativo:**

- 1) A pesar de que la ficha de toma de datos es muy completa tan sólo se puede acotar por título, URL y clasificación.
- 2) No se controla el vocabulario.
- 3) Las ayudas no son todo lo completas que debieran ser, pues no explican la significación de todos los datos que aparece en el formato de visualización completo ni los criterios de relevancia utilizados para ordenar los resultados de la búsqueda.

**LOOKSMART** (Looksmart International Limited)

<http://www.looksmart.com/>

**Contenido:** cobertura internacional con versiones locales en Australia y el Reino Unido.

**Recogida de Páginas:** existe un formulario de alta donde los interesados pueden sugerir nuevos recursos. En dicho formulario se demanda título, URL y descripción como datos obligatorios, y palabras-clave y fecha en la que el recurso fue puesto en la red como datos opcionales. Los datos deben de ser rellenados en el epígrafe clasificadorio en el que se piensa debe de estar enclavado el recurso. Esta información es revisada por un equipo de expertos de Looksmart quienes dan el alta definitiva.

**Búsqueda:** las búsquedas pueden realizarse o bien descendiendo por la clasificación temática o utilizando un formulario. El formulario es muy simple, pues tan sólo cuenta con una ventana donde han de teclearse los términos. En el formulario de búsqueda las palabras tecleadas son por defecto unidas con el operador AND. No se pueden utilizar ni operadores de búsqueda ni truncar ni localizar términos compuestos. No se puede acotar por campos. Existe la posibilidad de enlazar con el formulario para búsquedas avanzadas de Hotbot y lanzar la interrogación en este sistema.

**Resultados:** cuando se desciende por la clasificación lo que se obtiene son los recursos dados de alta en Looksmart. Si se utiliza el formulario de búsqueda, en primer lugar aparecerán los recursos existentes en Looksmart y a continuación las páginas obtenidas en Altavista. En el primer caso se obtiene el título con el enlace, la descripción y la categoría de la clasificación temática a la que pertenece con un enlace activo. En el segundo los datos proporcionados por Altavista. En el caso de Looksmart no se aplican criterios de relevancia en el orden de presentación de los recursos.

**Ayudas:** las ayudas son escasas en cuanto a la forma de realizar las búsquedas.

#### **Positivo:**

1. El alta de recursos se realiza a través de un formulario donde se solicita título, URL, clasificación, descripción, palabras clave y fecha de puesta en marcha del recurso.
2. Todos los recursos están clasificados siguiendo su índice temático.
3. Se recuperan recursos y no páginas.
4. Los recursos van acompañados de un resumen descriptivo del contenido.
5. Las ayudas son aceptables.

#### **Negativo:**

- 1) No se puede utilizar ningún operador de búsqueda ni truncar ni localizar términos compuestos.
- 2) No se puede acotar por campos.
- 3) No se controla el vocabulario.
- 4) No se puede seleccionar formatos.
- 5) No se aplica ningún criterio de relevancia en el orden de presentación de recursos.
- 6) No se pueden detectar las novedades.

### **OLE (Ole)**

<http://www.ole.es/>

**Contenido:** Web en español.

**Recogida de Páginas:** el método utilizado para localizar e incluir recursos es el alta por parte de los responsables a partir de un formulario. Antes de introducir los datos es necesario posicionarse en aquella categoría de la clasificación general en la que se piensa debe ubicarse el nuevo recursos. En el formulario se solicitan los siguientes datos: título, URL, país, idioma, ámbito, zona de influencia, tipo de servicio, descripción, palabras-clave, e-mail de contacto técnico y de contacto de contenido.

**Búsqueda:** se puede acceder a la información de dos modos: o descendiendo por el árbol de sectores o realizando una búsqueda desde cualquier punto de este mismo árbol utilizando un formulario desde donde se definen las búsquedas debiendo decidir si se truncan las palabras, si se enlazan con AND o con OR y en que campo se desea sean localizados los términos: título, URL, descripción, palabras clave, zona de influencia, sector, idioma, ámbito, tipo de servicio y país de origen. No se puede utilizar el operador NOT ni buscar términos compuestos ni usar paréntesis. En su opción "Temas" ofrece la posibilidad de visualizar por orden alfabético todas las categorías de su clasificación y en la de "Árbol" de desplegar la clasificación por apartados.

**Resultados:** no existe posibilidad de elección de formato de salida, apareciendo los recursos seleccionados agrupados por categorías y dentro de cada una de ellas por orden alfabético. De cada recurso se presenta el nombre con el enlace y la descripción. Se destacan las nuevas incorporaciones colocándolas por delante una "N".

**Ayudas:** en la ayuda se orienta muy someramente sin ejemplos sobre como se debe realizar una búsqueda.

**Positivo:**

- 1) El alta de recursos se realiza a través de un formulario donde se solicitan los siguientes datos: título, URL, país, idioma, ámbito, zona de influencia, tipo de servicio, descripción, palabras-clave, e-mails de contacto técnico y de contacto de contenido. El alta debe de hacerse desde el epígrafe clasificatorio adecuado.
- 2) Los recursos están organizados según una clasificación temática.
- 3) Se puede limitar por los siguientes campos: título, URL, descripción, palabras clave, zona de influencia, sector, idioma, ámbito, tipo de servicio y país de origen.
- 4) Se recuperan recursos de información y no páginas.
- 5) En la visualización de resultados se reproduce la descripción proporcionada por el responsable del recurso.
- 6) Las novedades pueden ser detectadas, pues las nuevas incorporaciones se destacan del resto durante 15 días a partir de su incorporación.

**Negativo:**

- 1) Las posibilidades de búsqueda son limitadas pues se puede truncar y utilizar los operadores AND y OR, pero no se puede utilizar el operador NOT ni el paréntesis ni localizar términos compuestos.
- 2) No se controla el vocabulario.
- 3) No se puede elegir formato de salida.
- 4) No se utilizan criterios de relevancia en la presentación de recursos.
- 5) Las ayudas no son muy completas y no presentan ejemplos de búsquedas.

**DONDE?** (Grupo Nuevas Tecnologías de la Información - Universidad Jaume I)

<http://donde.uji.es/>

**Contenido:** Web españoles. Cerca de 35.000 URL's.

**Recogida de Páginas:** los recursos se recogen a partir de un formulario de alta que demanda los siguientes datos: título, URL, tipo de recurso, datos institucionales (nombre, unidad o departamento y naturaleza de la organización responsable), descripción del contenido, localización geográfica (ciudad y comunidad autónoma), persona y e-mail del responsable del recurso.

**Búsqueda:** los recursos se encuentran agrupados por comunidades autónomas. Las búsquedas se pueden realizar en la totalidad de la base o en una comunidad autónoma determinada. Cuando se decide ir directamente a una comunidad autónoma es necesario visualizar por orden alfabético los recursos agrupados bajo esa categoría. En el caso de que se decida hacer una búsqueda en la totalidad de la base, se puede utilizar o un

formulario de una única ventana o bien otro más completo donde se puede acotar por los siguientes campos: título, URL, descripción, organización responsable, localidad y comunidad autónoma. Se permite el uso del asterisco para truncar, de las comillas para la localización de términos compuestos, y de los operadores lógicos AND y OR. No se puede utilizar ni el operador lógico NOT ni los paréntesis.

**Resultados:** se puede elegir entre un de formato de salida reducido y otro más completo donde aparece título, organización responsable, descripción, ciudad, comunidad autónoma y grado de relevancia. La visualización está limitada a 100 registros ordenados según su nivel de relevancia. En la presentación de páginas se utiliza un orden de relevancia. En los recursos recuperados aparece un resumen con el contenido del mismo. Los recursos más recientes tienen una etiqueta que indica su novedad. También existe un apartado denominado "Lo + nuevo", donde se agrupan las últimas entradas en orden cronológico inverso.

**Ayudas:** en las ayudas se informa sobre el funcionamiento en general del sistema y de la manera de realizar las búsquedas (aunque con muy pocos ejemplos). No se menciona los criterios de relevancia a la hora de la presentación de recursos.

**Positivo:**

- 1) El alta de recursos se realiza a través de un formulario donde se solicitan los siguientes datos: título, URL, tipo de recurso, datos institucionales, descripción del contenido, localización geográfica, persona y e-mail del responsable del recurso.
- 2) Se puede acotar por los siguientes campos: título, URL, descripción, tipo de organización responsable, localidad y comunidad autónoma.
- 3) Se recuperan unidades de información y no páginas.
- 4) Se puede elegir entre dos formatos de salida.
- 5) El formato de presentación completo presenta abundante información sobre el recurso en cuestión: título, institución responsable, ciudad, comunidad autónoma, resumen de contenido y grado de relevancia.
- 6) Las páginas recuperadas se presentan siguiendo un orden de relevancia
- 7) Los recursos más nuevos presentan una etiqueta que indica su novedad.

**Negativo:**

- 1) Los recursos no están organizados atendiendo a una clasificación temática.
- 2) Las posibilidades de búsqueda no son completas, pues aunque se puede truncar, buscar términos compuestos y los operadores

- AND y OR, no se pueden usar ni el operador NOT ni los paréntesis.
- 3) No se controla el vocabulario.
  - 4) Las ayudas no son muy completas, pues apenas se presentan ejemplos de búsquedas y no se mencionan los criterios de relevancia a la hora de presentación de recursos.

**OZU.com** (OzuCom S.L.)

<http://www.ozu.com/>

**Contenido:** Web españoles.

**Recogida de Recursos:** a través de un formulario donde debe consignarse la URL, el título y la descripción de la página junto con el e-mail en el caso de que la inscripción haya sido hecha por el creador o el administrador de la misma. Los recursos en OZU.com están organizados siguiendo una clasificación temática. El alta debe hacerse desde el epígrafe donde se piensa pueda estar incluido un recurso dado.

**Búsqueda:** la búsqueda puede realizarse o bien descendiendo por la clasificación temática hasta encontrar el epígrafe deseado o bien utilizando uno de los dos formularios de búsqueda que existen. Uno de ellos tan sólo tiene una única ventana donde deben escribirse los términos. El otro formulario es más completo y permite acotar por campos (URL, título y descripción), seleccionar un epígrafe determinado de la clasificación general, elegir entre utilizar los operadores AND y OR y truncar las palabras si así se desea. No se puede utilizar el operador NOT, ni los paréntesis ni localizar términos compuestos. Cuando los términos de busca son muy generales el sistema presenta los epígrafes de la clasificación y obliga a restringir a uno de ellos.

**Resultados:** no se puede elegir el formato de salida. Cuando se lanza una pregunta, primero se obtienen las secciones de la clasificación que concuerdan con las palabras buscadas, a continuación aparecen los recursos (título, descripción y URL) ordenados por secciones de la clasificación. Dentro de cada epígrafe clasificatorio los recursos se presentan por orden alfabético. A las páginas que han sido añadidas recientemente se les añade una etiqueta que así lo indica. Existe un apartado donde se presentan agrupados por epígrafes los recursos dados de alta recientemente.

**Ayuda:** se da información sobre el funcionamiento de OZU.com, y se detallan las posibilidades de búsqueda sin incluir ejemplos.

**Otros:** OZUEXPRESS es un servicio de búsqueda a la carta que envía periódicamente nuevas páginas relacionadas con el tema de interés

expresado en el formulario existente en esta sección. Se puede elegir entre la periodicidad del envío de las novedades (mensual, semanal o diario).

**Positivo:**

- 1) El alta de los recursos se hace a través de un formulario donde se solicita la URL, el título y la descripción de la página, una vez situado en el epígrafe clasificatorio adecuado.
- 2) Los recursos están organizados atendiendo a una clasificación temática.
- 3) Se puede acotar por los siguientes campos: URL, título y descripción.
- 4) Se recuperan recursos y no páginas.
- 5) De los recursos recuperados se presenta las descripciones de los mismos.
- 6) Las nuevas incorporaciones se destacan con una etiqueta.
- 7) Las ayudas son bastante completas, aunque no presentan ejemplos de búsqueda.
- 8) Existe un servicio de búsqueda a la carta con envío periódico de resultados.

**Negativo:**

- 1) Las posibilidades de búsqueda son deficientes, pues no se pueden localizar términos compuestos ni utilizar el operador NOT ni los paréntesis.
- 2) No se puede elegir formato de salida.
- 3) No existen criterios de relevancia en la presentación.

**ELCANO**

<http://www.elcano.es/>

**Contenido:** Web principalmente del ámbito hispano.

**Recogida de Páginas:** las páginas se recogen a partir de un formulario donde se solicitan los siguientes datos: URL, título, tipo de recurso (lista de correo, publicación electrónica, etc.), apartado del índice temático donde debe insertarse, naturaleza del propietario (empresa comercial, institución académica, etc.), idioma, descripción y e-mail del responsable. Todos los registros son revisados antes de ser dados definitivamente de alta en la base de datos.

**Búsqueda:** las búsquedas pueden realizarse o bien entrando por la clasificación temática que está dividida en dos partes diferenciadas, "Áreas de Interés" y "Profesionales y Colectivos", hasta encontrar el epígrafe

adecuado, o bien a través de un formulario donde se deben escribir los términos de la búsqueda y elegir el idioma y naturaleza del propietario de las páginas. No se pueden utilizar ni operadores ni paréntesis. Las palabras son interpretadas como cadenas de caracteres y cuando se teclea más de una como frase.

**Resultados:** No se puede elegir el formato de salida. Los resultados aparecen ordenados por fecha de alta en el sistema, en primer lugar los más recientes. De cada recurso se presenta el título con el enlace, la descripción, el idioma, el tipo de recurso, la fecha de alta, el e-mail del responsable y el nombre de la persona que ha proporcionado la información. A todos los recursos se les otorga una calificación. **Ayudas:** No existe ningún tipo de ayuda.

**Otros:** Existe un apartado de novedades donde aparecen todos los recursos dados de alta más recientemente, ordenados por fecha de alta.

**Positivo:**

- 1) El alta se realiza a partir de un formulario donde se solicitan los siguientes datos: URL, título, tipo de recurso, epígrafe clasificatorio, tipo de propietario, idioma, descripción y e-mail del responsable.
- 2) Existencia de una clasificación temática.
- 3) Se recuperan recursos de información y no páginas.
- 4) De cada recurso recuperado se presenta el título con el enlace, la descripción, el idioma, el tipo de recurso, la fecha de alta, el e-mail del responsable y el nombre de la persona que ha proporcionado la información.
- 5) Todos los recursos están valorados.
- 6) Los recursos se presentan por fecha de alta, lo que en todo momento permite conocer los más recientes.

**Negativo:**

- 1) Las posibilidades de búsqueda son bastante deficientes, ya que no se pueden utilizar ni operadores lógicos ni paréntesis y las palabras siempre son interpretadas como cadenas de caracteres. Se pueden localizar términos compuestos.
- 2) Exclusivamente se puede acotar por idioma y por tipo de propietario.
- 3) No se controla el vocabulario.
- 4) No se puede elegir el formato de salida.
- 5) No se aplican criterios de relevancia en la presentación de recursos.
- 6) No existe ningún tipo de ayuda.

**OZU.es (AdverNet, S.L.)**

<http://www.ozu.es/>

**Contenido:** Web españoles.

**Recogida de Recursos:** a través de un formulario donde debe consignarse la URL, el título, la descripción de la página y la sección de la clasificación temática de OZU donde se piensa debe estar ubicada. También se solicita el nombre y el e-mail de la persona que realiza el alta.

**Búsqueda:** la búsqueda puede realizarse o bien descendiendo por la clasificación temática hasta encontrar el epígrafe deseado o bien utilizando un formulario donde se elige entre localizar los términos de búsqueda en el documento completo o en alguno de los siguientes campos: título, descripción, URL o e-mail. También se puede realizar la búsqueda en el interior de cada una de las categorías del índice, en una comunidad autónoma o en la totalidad de OZU.

El operador utilizado por defecto en el formulario de búsqueda es el OR. Se puede utilizar el operador AND, pero no así el NOT. Tampoco se puede hacer uso de los paréntesis. Las palabras son interpretadas por el sistema como cadenas de caracteres y no como palabras delimitadas. Se permite el uso de comillas para la localización de términos compuestos.

**Resultados:** no se indica el número total de respuestas. No se puede elegir el formato de salida. Los resultados se presentan agrupados por epígrafes clasificatorios y dentro de cada uno de ellos por orden alfabético. De cada recurso se presenta el título con el enlace, la descripción y el e-mail del responsable del mismo.

**Ayuda:** no detalla los operadores que pueden ser utilizados.

**Positivo:**

- 1) El alta se realiza a través de un cuestionario donde se solicitan la URL, el título, la descripción de la página y la sección de la clasificación temática.
- 2) Existe una clasificación temática.
- 3) Se puede acotar por los siguientes campos: URL, título, descripción, y comunidad autónoma.
- 4) Se recuperan recursos de información y no páginas.
- 5) Los recursos seleccionados aparecen acompañados de una descripción.

**Negativo:**

- 1) Las posibilidades de búsqueda son limitadas pues no se puede utilizar ni el operador NOT ni usar los paréntesis. Además el sistema interpreta lo escrito como cadenas de caracteres y no como palabra completa.

- 2) No se controla el vocabulario.
- 3) No se indica el resultado total de las búsquedas.
- 4) No se puede elegir formatos de salida.
- 5) No existe ningún orden de relevancia para establecer el orden de presentación de los resultados.
- 6) No se detectan las novedades.
- 7) Las ayudas son muy exigidas.

**BIWE** (Cesat)

<http://biwe.cesat.es/>

**Contenido:** Web españoles.

**Recogida de Páginas:** a través de los propios usuarios y responsables de los Web, quienes una vez situados en la categoría pertinente del índice temático rellenan un formulario con los siguientes datos del web a dar de alta: título, URL, e-mail, breve descripción, localización, categoría donde se da de alta y palabras-clave. La información introducida por este sistema es procesada y verificada por el sistema informático.

**Búsqueda:** la búsqueda se realiza a través de dos procedimientos: o bien descendiendo por las diferentes categorías en las que están distribuidos los recursos o bien utilizando uno de los dos formularios de búsqueda que presenta. El más sencillo consiste en una única ventana donde se escriben los criterios de búsqueda realizándose por defecto la unión entre ellos. En el formulario más completo se pueden utilizar exclusivamente los operadores AND y OR; deben limitarse a los campos de búsqueda existentes: URL, título, descripción, palabras-clave; y no se permite el uso de términos compuestos ni paréntesis ni se puede truncar. Una vez obtenido el resultado de una búsqueda se puede acotar más utilizando nuevos términos.

**Resultados:** no existe posibilidad de elección de formato de salida, ya que siempre aparecen el título y la descripción, pero si se puede optar entre presentar los recursos agrupados por categorías o no y elegir el orden de presentación entre las siguientes opciones: "los más accedidos", "los más nuevos" o "por orden alfabético".

**Ayudas:** en las ayudas se aconseja sobre como realizar una búsqueda utilizando ejemplos. También se dan explicaciones sobre la presentación de los recursos.

**Otros:** Biwexpres.

**Positivo:**

- 1) El alta se hace a través de un formulario donde se solicita: título, URL, e-mail, descripción, localización, categoría de la clasificación y palabras-clave.
- 2) Todos los recursos se encuentran ordenados siguiendo una clasificación temática.
- 3) Se puede acotar por los siguientes campos: URL, descripción y palabras-clave.
- 4) Se recuperan unidades de información y no páginas.
- 5) Los recursos recuperados vienen acompañados por una descripción.
- 6) Se pueden detectar las novedades al elegir la opción "los más nuevos" entre las diferentes posibilidades de presentación de resultados.
- 7) Las ayudas son bastante completas.

**Negativo:**

- 1) Las posibilidades de búsqueda son muy limitadas, pues no se puede utilizar ni el operador NOT ni los paréntesis ni localizar términos compuestos ni truncar.
- 2) No se controla el vocabulario.
- 3) No se pueden seleccionar formatos de salida.
- 4) La ordenación de resultados no corresponde a ningún criterio de relevancia.

**TROVATOR** (Combios)

<http://trovator.combios.es/>

**Contenido:** Web en español: 400.000 páginas hispanas y 15.000 Web clasificados.

**Recogida de Páginas:** aunque Trovator cuenta con un robot capaz de moverse y localizar nuevas páginas autónomamente, existe un formulario que puede ser utilizado para dar de alta a nuevos recursos en el propio Trovator y en Señas, directorio de recursos clasificados. En este formulario se solicitan los siguientes datos de las páginas a introducir: título, URL, epígrafes de la clasificación de Señas, descripción, país de origen y e-mail.

**Búsqueda:** las búsquedas se realizan descendiendo por la clasificación de Señas o introduciendo los términos de búsqueda en un formulario que permite acotar por dominio. Se puede elegir entre introducir los términos directamente sin ningún vínculo entre ellos o realizar una búsqueda avanzada y utilizar todos los operadores permitidos: AND, OR, NOT, NEAR, además del asterisco para truncar. También se pueden usar paréntesis. Tan sólo se puede acotar por países.

**Resultados:** no se pueden elegir formatos. En el caso de Trovator aparecen las páginas ordenadas según un orden de relevancia, presentándose tan sólo la URL con el enlace y en la mayoría de los casos las primeras palabras de la página a modo de descripción. En el caso de Señas aparecen los títulos y la descripción de los recursos también sin ningún orden aparente.

**Ayudas:** explicación breve pero clara de cómo realizar una búsqueda. No dice nada sobre los criterios de relevancia aplicados a la presentación de las páginas.

**Otros:** Existe una sección de nuevos donde aparecen sin ningún orden las páginas introducidas en Trovator en la semana en curso.

**Positivo:**

- 1) Sistema mixto de recogida de nuevas páginas a través de un robot o de un formulario en el que se solicita: título, URL, epígrafes de la clasificación de Señas, descripción, país de origen y e-mail.
- 2) Las posibilidades de búsqueda son muy completas, pues se pueden utilizar todos los operadores booleanos y los paréntesis, se puede truncar y buscar términos compuestos.
- 3) Los resultados se presentan ordenados según ciertos criterios de relevancia.
- 4) Las ayudas son suficientes en el caso de la búsqueda, pero no explican los criterios de relevancia aplicados en la presentación de los resultados.

**Negativo:**

- 1) Los recursos están clasificados por temas sólo parcialmente.
- 2) Sólo se puede acotar por país.
- 3) No se controla el vocabulario.
- 4) Se recuperan principalmente páginas.
- 5) No se puede elegir formatos de salida.
- 6) No todas las páginas aparecen con una descripción propia.
- 7) Las páginas recuperadas no pueden ser agrupadas por servidores.
- 8) Es muy difícil la detección de novedades, pues aunque existe un apartado que agrupa las altas de la semana en curso, no se pueden realizar búsquedas.

**TARANTULA** (Intercomputers, S.A. de C.V.)

<http://www.tarantula.com.mx/>

**Contenido:** Web en español.

**Recogida de Páginas:** a través de un formulario de alta donde tan sólo es necesario escribir la URL de la página, el nombre del responsable de la misma y su e-mail. A continuación Tarántula inspeccionará las páginas automáticamente.

**Búsqueda:** las búsquedas se realizan a partir de un formulario con una única ventana. Las palabras tecleadas sobre él son interpretadas como cadenas de caracteres. El sistema utiliza el operador AND por defecto, no reconociendo ni el OR ni el NOT. No se puede utilizar paréntesis. No se pueden localizar términos compuestos. No se puede acotar por campos.

**Resultados:** no se puede elegir el formato de salida. Las páginas se presentan por orden de relevancia, apareciendo la siguiente información de las mismas: título con enlace, primeras palabras a modo de descripción, e-mail del responsable, tamaño en bits y fecha de alta. No se señalan los recursos nuevos.

**Ayuda:** no se da apenas información del funcionamiento del sistema en general ni de como realizar las búsquedas.

**Positivo:**

**Negativo:**

- 1) No existe clasificación temática.
- 2) Las posibilidades de búsqueda son incompletas, pues no se puede utilizar ni los operadores OR y NOT ni los paréntesis ni localizar términos compuestos.
- 3) No se puede acotar por campos.
- 4) No se controla el vocabulario.
- 5) Se recuperan páginas y no recursos de información que no pueden ser agrupadas por servidores.
- 6) No se pueden elegir formatos de salida.
- 7) Las páginas recuperadas se acompañan de las primeras líneas a modo de descripción.
- 8) No existen criterios para establecer el orden de presentación de páginas recuperadas
- 9) No se distinguen las páginas nuevas.
- 10) Prácticamente no existen ayudas.

**HISPAVISTA** (Hispavista)

<http://www.hispavista.com/>

**Contenido:** Web en español.

**Recogida de Páginas:** existe un formulario para dar de alta a nuevas páginas, en el que hay que facilitar los siguientes datos: título, URL, categoría de la clasificación a elegir entre los epígrafes sugeridos), descripción y país de la página sugerida junto con el nombre y el e-mail de una persona de contacto.

**Búsqueda:** la búsqueda puede realizarse o bien descendiendo por la clasificación temática hasta encontrar el epígrafe deseado o bien utilizando uno de los dos formularios de búsqueda que existen. Uno de ellos tan sólo tiene una única ventana donde deben escribirse los términos. El otro formulario, difícil de encontrar, es un poco más completo y da la posibilidad de truncar y de buscar términos compuestos, aunque el sistema no es capaz de localizar términos compuestos unidos por preposiciones desde la opción palabra exacta. Las palabras por defecto se relacionan con el operador AND. El operador OR lo reconoce como AND y no funciona el NOT ni el paréntesis. No se puede acotar por campos.

**Resultados:** no se puede elegir el formato de salida. Cuando se lanza una pregunta, las páginas seleccionadas se presentan bajo su epígrafe clasificatorio, reproduciéndose el título y la descripción. Dentro de cada epígrafe clasificatorio los recursos se listan por orden alfabético. Existe un apartado donde se presentan agrupados por epígrafes los recursos dados de alta recientemente.

**Ayuda:** se da información muy general sobre el funcionamiento del sistema no detallándose adecuadamente las posibilidades de búsqueda. No se incluyen ejemplos.

#### **Positivo:**

- 1) Las altas se producen a través de un formulario donde se solicitan los siguientes datos: título, URL, categoría de la clasificación, descripción y país con el nombre y el e-mail de una persona de contacto.
- 2) Todos los recursos están ordenados siguiendo una clasificación temática.
- 3) Se recuperan recursos de información y no páginas aisladas.
- 4) Los recursos recuperados van acompañados con una descripción de los mismos.
- 5) Se pueden detectar las novedades en un apartado especial.

#### **Negativo:**

- 1) Las posibilidades de búsqueda son incompletas, pues no funcionan los operadores OR ni NOT y no se pueden utilizar paréntesis.
- 2) No se puede acotar por campos.
- 3) No existe control de vocabulario.

- 4) No se pueden elegir formatos de salida.
- 5) No existen criterios de presentación.
- 6) Las ayudas son muy exigidas.

## **SOL (Sol)**

<http://www.sol.es/>

**Contenido:** Web españolas.

**Recogida de Páginas:** las páginas son recogidas por un robot que es capaz de indizarlas y de seguir los demás links. Las altas se realizan a partir de un formulario donde únicamente se solicita la dirección de la página a dar de alta.

**Búsqueda:** presenta un único formulario de búsqueda. Se pueden utilizar los operadores booleanos (AND, OR, NOT), las comillas para localizar términos compuestos y el asterisco para truncar palabras. No se puede utilizar paréntesis ni acotar por campos.

**Resultados:** se puede elegir entre visualizar el formato completo y otro simplificado. En el completo aparece el título de la página, las primeras palabras de la misma a modo de descripción, la URL, el tamaño de la página en bits y la fecha en que se han introducido en el sistema. No se destacan las páginas nuevas. Se pueden mantener abiertas simultáneamente la página de resultados y una de las recuperadas.

**Ayudas:** se explica cuáles son los operadores que se pueden utilizar y la forma de realizar búsquedas presentando ejemplos. No se habla de los criterios de ordenación de las páginas una vez realizadas una búsqueda.

### **Positivo:**

- 1) Se puede elegir entre dos formatos de visualización.
- 2) Se puede mantener abiertas simultáneamente la página de resultados y una de las páginas recuperadas.
- 3) Las ayudas son bastante completas a la hora de explicar el modo de realización de las búsquedas.

### **Negativo:**

- 1) No existe clasificación temática.
- 2) Las posibilidades de búsqueda son incompletas, pues no se pueden utilizar los paréntesis.
- 3) No se puede acotar por campos.
- 4) No existe control de vocabulario.

- 5) Se recuperan páginas y no recursos de información que no pueden ser agrupadas por servidores.
- 6) En las páginas recuperadas aparecen las primeras líneas del texto a modo de descripción.
- 7) No existen criterios de relevancia en la presentación de resultados.
- 8) No se pueden detectar las novedades.

## 2. ANALISIS DE RESULTADOS

Una vez estudiadas las características documentales de cada uno de los buscadores seleccionados y a partir de las mismas se pueden hacer las siguientes observaciones.

### a. TIPOS DE BUSCADORES

De los veinte buscadores seleccionados, nueve son motores de búsqueda y once índices temáticos. De los nueve motores, seis son internacionales: Altavista, Excite, Lycos, WebCrawler, Hotbot Infoseek; y tres reúnen URLs en español: Trovator, Tarántula y Sol. Al contrario, de los once índices temáticos, siete contienen direcciones en español: Ole, Donde?, Ozu.com, Elcano, Ozu.es, Biwe e Hispavista; los cuatro restantes tienen una cobertura internacional.(Yahoo!, Magellan, Galaxy y Looksmart.

### b. RECOGIDA Y ANALISIS

#### - Buscadores Internacionales

De los seis motores de búsqueda internacionales tan sólo tres (Altavista, Infoseek y Hotbot) reconocen y utilizan las etiquetas Meta de las páginas recogidas. Los otros tres (Excite, Lycos y WebCrawler) no las reconocen o las ignoran.

Los cuatro buscadores restantes son índices que recogen la información de los recursos a través de un formulario de alta, tres de ellos (Yahoo!, Galaxy y Looksmart) solicitan todos los datos considerados imprescindibles (Título, URL, clasificación y descripción). Para los datos considerados recomendables estos son los resultados: palabras-clave (1/4), localización (1/4), idioma (1/4), tipo de recurso (1/4) y propietario de las páginas (0/4). En Magellan tan sólo se solicita la URL del recurso a dar de alta y un equipo interno extrae la información, aunque no se ha podido averiguar los datos extraídos. El que más información recoge es Galaxy, que de los nueve datos solicitados, recoge información de siete.

#### - Buscadores en español

De los siete índices que recogen recursos en español, excepto Donde?, que no clasifica sus direcciones, el resto recogen de cada uno de los recursos los datos considerados imprescindibles: título, URL, clasificación y descripción. Pero ninguno de ellos recoge todos los datos recomendables (palabras-clave, localización, idioma, tipo de información y tipo de propietario): Ole tres, Elcano y Donde? y Biwe dos e Hispavista uno. De los tres motores de búsqueda (Trovator, Tarántula y Sol) no se ha podido averiguar si reconocen y analizan el contenido de las etiquetas Meta.

### c. BUSQUEDA

#### - Buscadores internacionales

- Formularios de búsqueda. Excepto WebCrawler, Magellan y Looksmart que no presentan nada más que un formulario simple para realizar las búsquedas, el resto de los buscadores dan la posibilidad de elegir entre uno simple con una única ventana y otro más completo donde plasmar estrategias de búsqueda más complejas. Sólo en el caso de Altavista las búsquedas avanzadas se realizan en un formulario de una única ventana.
- Herramientas de búsqueda. Casi todos los buscadores permiten utilizar la totalidad de los operadores de booleanos (AND, OR y NOT) y los paréntesis, truncar y buscar términos compuestos. Las excepciones son las siguientes: Yahoo, Looksmart y Galaxy no permiten utilizar el paréntesis; Infoseek, Lycos, WebCrawler, Magellan y Looksmart no permiten truncar. Además en Looksmart no se puede ni utilizar los operadores OR y NOT ni localizar términos compuestos.
- Clasificación temática. Tanto los índices como los motores de búsqueda presentan una clasificación temática a través de la cual se pueden localizar direcciones. La diferencia entre los índices y los motores es que en los primeros todas las direcciones están clasificadas, mientras que en los segundos solamente están clasificados una parte de sus recursos.
- Campos de búsqueda. De los índices se puede decir que es Yahoo! el único que permite acotar por los tres campos considerados imprescindibles (título, URL y descripción), además de por idioma. Galaxy admite que se acote por título y URL, y Magellan y Looksmart no dan la oportunidad de localizar direcciones utilizando campos determinados, a pesar de que en sus formularios de toma de datos solicitan los dos: título, URL y descripción de las direcciones a dar de alta. Galaxy pide además la localización, el idioma y el tipo de información y Looksmart las palabras clave.
- De los motores que son capaces de reconocer y extraer la información de las etiquetas Meta: Altavista permite seleccionar páginas por título, URL, localización e idioma; Infoseek por título, URL y localización; y Hotbot exclusivamente por título y localización. Del resto de los motores que

afirman no utilizar las etiquetas Meta: Excite permite acotar exclusivamente por idioma, Lycos por título, URL e idioma; y WebCrawler no permite acotar.

- Control del vocabulario: ninguno de los buscadores controla el vocabulario, salvo en cierta forma Excite y Magellan. El sistema que utilizan ambos para llevar a cabo dicho control, no está lo suficientemente explicado.
- Detección de novedades: de los diez buscadores estudiados en tan sólo cuatro se pueden detectar las nuevas incorporaciones: Yahoo! (acotación por fecha de alta del recurso), Infoseek (registro de salida con fecha de alta en el sistema), Hotbot (fecha de creación y modificación de páginas) y Galaxy (fichero aparte).

### Buscadores en español

- Formularios de búsqueda. De los diez buscadores estudiados, todos presentan una pantalla de búsqueda simple, y seis de ellos (Ole, Donde?, Ozu.com, Ozu.es, Biwe e Hispavista) añaden otra más completa para realizar estrategias de búsqueda más complejas.
- Herramientas de búsqueda. Menos Elcano, el resto de buscadores pueden interseccar los términos utilizando el operador booleano AND. Sin embargo, el grupo que da la oportunidad de utilizar el OR es menor, pues Elcano, Tarántula e Hispavista no permiten su uso. El NOT y los paréntesis tan sólo pueden ser usados en Trovator y Sol. En todos los buscadores, salvo en Biwe, se puede truncar. Donde?, Ozu.es, Trovator y Sol dan la oportunidad de localizar términos compuestos.
- Clasificación temática. De los diez buscadores estudiados, siete de ellos, es decir la totalidad de los índices, dan la posibilidad de buscar a través de la clasificación temática utilizada para la ordenación de sus recursos. En el caso de los motores de búsqueda tan sólo Trovator clasifica parcialmente sus recursos y da la oportunidad de consulta a través de un índice.
- Campos de búsqueda: de los tres motores de búsqueda hay que señalar que tan sólo Trovator permite la interrogación por localización como único campo. En los índices las posibilidades de acotar por campos son mayores. Ole es el más completo en este aspecto, pues permite preguntar por título, URL, descripción, palabras-clave, localización e idioma. A continuación le sigue Donde? que ofrece la posibilidad de interrogar por los campos considerados como imprescindibles (título, URL y descripción), además de por localización y tipo de propietario. En Ozu.com y Ozu.es se puede preguntar igualmente por los campos considerados imprescindibles, añadiendo Ozu.es la posibilidad de acotar por localización. En Biwe se pueden utilizar los campos URL, descripción y palabras-clave. Finalmente los índices más pobres en este apartado son Elcano (idioma y tipo de propietario) e Hispavista (ninguno).

- Control de vocabulario: ninguno de los buscadores controla el vocabulario en la interrogación.
- Detección de novedades: de los tres motores de búsqueda (Trovator, Tarántula y Sol) en ninguno se puede detectar las nuevas incorporaciones. Sin embargo, la situación es la contraria para los índices, pues, salvo en Ozú.es, en el resto se pueden detectar a través de varios procedimientos: con etiquetas especiales (Ole, Donde?, Ozu.com), poniendo la fecha de alta en los registros recuperados (Elcano) y con un fichero aparte (Biwe e Hispavista)

Si se comparan de manera global los buscadores internacionales con los buscadores en español, se llega a la conclusión de que los internacionales son más completos desde el punto de vista de la búsqueda, pues alcanzan un mayor grado en los apartados de pantallas, herramientas de búsqueda, clasificación de recursos y control de vocabulario. En el caso de la búsqueda por campos las posiciones se aproximan más. En el único apartado en que los buscadores españoles aventajan a los internacionales es en la detección de novedades.

### a. RESULTADOS

- Buscadores internacionales

Tan sólo pueden elegirse formatos de salida en cuatro de los diez buscadores estudiados: Excite, WebCrawler, Hotbot y Galaxy. Sin embargo, en los diez las páginas y los recursos de información vienen acompañados con una descripción de su contenido, en el caso de los motores de búsqueda representados por las primeras palabras aparecidas en las páginas o por el contenido de las etiquetas Meta en el caso de que existan. Todos los buscadores, tanto índices como motores, ordenan los registros de salida en función de unos criterios de relevancia. El punto más negativo de este apartado tiene que ver con la posibilidad de agrupación de las páginas recuperadas en los seis motores, pues tan sólo Excite y Lycos cuentan con esta opción.

- Buscadores en español

En el caso de los buscadores en español, se puede decir que en tan sólo en dos (Donde? y Sol) se puede elegir entre formatos de salida; que todos presentan una descripción de contenido de las páginas y los recursos gestionados; que únicamente en tres (Donde?, Trovator y Tarántula) se utilizan criterios de relevancia en el orden de presentación de resultados; y por último que ninguno de los motores de búsqueda permite agrupar las páginas recuperadas por servidores.

## BIBLIOGRAFIA

CODINA, Lluís. "Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte I)". *Information World en Español*, 1997, Vol. 6, nº 5, p. 22-27.

CODINA, Lluís. "Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte II)". *Information World en Español*, 1997, Vol. 6, nº 6, p. 18-26.

DÍEZ FERREIRA, Miguel Ángel. "Los 20 mejores buscadores". *World 9: Suplemento de Pcwold*, 1997, nº 137, pp. 34-44.

MARCOS MORA, Mari Carmen. "Motores de recuperación de información: un análisis comparativo (parte I)". *Information World en Español*, 1998, Vol. 7, nº 1-2, pp. 18-22.

MARCOS MORA, Mari Carmen. "Motores de recuperación de información: un análisis comparativo (parte II)". *Information World en Español*, 1998, Vol. 7, nº 3, pp. 13-20.

SENSO, J. A. "Herramientas para realizar búsquedas en Internet: una revisión". *Information World en Español*, 1998, Vol. 7, nº 1-2, pp. 24-25.

## Análisis del buscador múltiple *Copernic 2001*

**Franganillo, J. & Figuerola, T. M. (2001).**

**Biblioteconomía y Documentación.**

<http://www.ub.es/biblio/bid/06frang2.htm>

En los últimos años han aparecido varios buscadores múltiples en forma de software de gestión con cierto grado de sofisticación, como es el caso del programa que nos ocupa, que incorporan una serie de procesos automatizados: programación, almacenamiento y seguimiento de las consultas, y comprobación de la accesibilidad de las direcciones. En un intento de designar su propia especie, *Copernic* se presenta a sí mismo con la apelación poco modesta de *agente inteligente*, vanidad que desgraciadamente se ha extendido mucho. La atribución del adjetivo *inteligente* resulta totalmente gratuita porque el programa no demuestra ninguna capacidad que la justifique. Por lo tanto, es más acertado ceñirse a la esencia de estos programas y evitar los nombres imprecisos y cercanos a la ciencia-ficción. Proponemos, en consecuencia, tratarlos sencillamente de *agentes buscadores*.

**Presentación:** *Copernic*, producto de la empresa canadiense Copernic Technologies, es un buscador múltiple que transfiere una ecuación de búsqueda a un conjunto de buscadores de manera simultánea, recupera las referencias pertinentes y las ordena según el grado medio de relevancia obtenido de cada uno de los buscadores. *Copernic Pro*, la edición comercial más completa, ofrece, además, otras prestaciones interesantes: actualización de búsquedas programadas, servicio de alerta y verificador ortográfico de las búsquedas.

**Recuperación de la información:** El cuadro de diálogo permite escoger entre buscar por intersección (“Buscar todas las palabras”), por unión (“Buscar una de las palabras”), o por frase (“Buscar la expresión exacta”)

**Eliminación de duplicados:** La documentación de *Copernic* sólo menciona que el programa hace esta función, tal vaguedad demuestra que el mecanismo no funciona tan bien como sería deseable

**Validación de los documentos:** La función de validación identifica las direcciones inaccesibles (porque no existen o porque no se puede establecer la conexión en el tiempo de espera prefijado) y da la opción de eliminarlas

**Descarga de los documentos:** *Copernic* permite almacenar los documentos hallados para poder explorarlos después, sin conexión.

**Depuración de las búsquedas:** Una vez lanzada la consulta con el lenguaje de interrogación básico y obtenidos los resultados, es muy probable que convenga acotar la necesidad informativa porque las limitaciones en la búsqueda inicial pueden haber producido un resultado bastante ruidoso. A través de un lenguaje

de interrogación más sofisticado, la función “Refinar” permite filtrar un conjunto de recursos.

Presentación y ordenación de los resultados: En el historial de búsquedas se muestra un listado de las peticiones hechas y que *Copernic* almacena automáticamente. De cada una presenta la ecuación (sin refinar), el máximo de resultados (por buscador y en total), la modalidad de consulta (intersección, unión o frase), la fecha, la hora de creación y la última actualización, el número de referencias recuperadas, la categoría en la que se ha ejecutado y un campo de observaciones. El usuario puede configurar esta presentación de acuerdo con sus necesidades y escoger las columnas de datos y el orden de colocación. Las búsquedas aparecen en orden cronológico, pero pueden ordenarse bajo otros criterios.

Asignación de relevancia: *Copernic* no efectúa búsquedas, sino que las delega a otros buscadores, cada uno con criterios de ponderación diferentes, y calcula una media porcentual de los índices de relevancia obtenidos para cada referencia. El acierto en el cálculo de relevancia por parte de cada buscador activado depende de la calidad de su sistema de recuperación; unos se basan en los metadatos, otros se fijan en la posición de las palabras dentro del documento, otros valoran la cantidad de enlaces procedentes de otras páginas, y los hay que valoran la proximidad de las palabras o la frecuencia de aparición. La eficacia de esta ponderación puede cuestionarse ya que el orden no siempre es aceptable. Habitualmente, los recursos con una puntuación de relevancia mayor son pertinentes.

Exportación de los resultados: Mediante la función de exportación de los resultados, se pueden enviar informes o los resultados de las búsquedas por correo electrónico.

Servicio de alerta: *Copernic* ofrece un agente de seguimiento para realizar la actualización de las búsquedas de manera autónoma y notificar al usuario sobre la aparición de nuevos documentos de su interés o indicar los cambios que se hayan producido. Es posible programar el agente para que envíe por correo electrónico a una o varias direcciones los informes con los nuevos documentos hallados.

Traducción de documentos: *Copernic* ofrece una herramienta de traducción de documentos a través del servicio en línea *Gist-In-Time* (<http://translate.copernic.com:8090/>).