

EVALUACIÓN DE HERRAMIENTAS DE BÚSQUEDA DE INFORMACIÓN EN INTERNET

Por : **Alonso Estrada Cuzcano y Noelia Rodríguez Maniega**
Correo electrónico: **aestrada@udep.edu.pe**
Licenciados en Bibliotecología

Resumen

El conjunto de herramientas, que han aparecido con el desarrollo de Internet y que ayudan a la recuperación organizada de la información, hacen más difícil seleccionar las que mejor se adecuan a las necesidades de recuperación de información de cada persona.

El trabajo presentado reúne la evaluación de cuatro de las principales herramientas para la recuperación de la información en Internet. El investigador hizo búsquedas en cada una de estas herramientas para analizar los resultados, ver las características y potencialidades de cada motor de búsqueda que comparte con nosotros en estas páginas .

1. Introducción

Con el gran desarrollo de Internet ha sido necesaria la creación de herramientas que ayuden a la recuperación organizada de la información. Sin embargo, en la actualidad son muchas las herramientas que han aparecido y por lo tanto se hace más difícil seleccionar cuales son aquellas que se adecuan mejor a las necesidades de recuperación de información de cada persona.

Tanto los motores de búsqueda de primera y segunda generación, así como los directorios temáticos tienen todos su propio algoritmo de búsqueda y generan diferentes respuestas. No se dispone de procedimientos de búsqueda uniformes o normalizados, por lo cual es imposible brindar respuestas a una

solicitud de búsqueda con las mismas estrategias. Así, la tarea de buscar implica también encontrar el motor de búsqueda más satisfactorio o combinar varios.

Los sistemas de recuperación de información por Internet poseen una gran flexibilidad en el tratamiento de los datos y la multiplicidad de productores determinan una gran redundancia y a la vez una dispersión de la información distribuida; cada buscador indiza solo una parte del universo que contiene Internet. Además, existe una gran diversidad de soportes que causan una gran heterogeneidad en la estructura de los documentos (videos, gráficos, multimedia, etc.), mucha información en Internet está caracterizada por el dinamismo y la volatilidad. El dinamismo se refiere a los continuos cambios de contenido de muchos de los documentos de Internet y la volatilidad, a los cambios de destino de un mismo documento.

El trabajo incluye una metodología para realizar las búsquedas y un posterior análisis de los resultados, estudia las características de cada buscador, así se denominará en adelante tanto a los motores de búsqueda como a los directorios; se realizarán estrategias de búsqueda y finalmente se analizarán la exhaustividad, pertinencia y relevancia a partir de los resultados que se obtiene de cada buscador.

Por tanto se pretende hacer una pequeña aportación con el análisis de cuatro de estas herramientas para sacar cuales son sus diferencias y características.

2. Metodología

La **metodología** que se ha seguido para desarrollar esta evaluación y descripción de buscadores es la que a continuación se detalla:

1. Recogida de consultas a partir de preguntas realizadas al colectivo universitario de la Universidad Carlos III de Madrid. Fueron seleccionados dos representantes de cada tipo de personal existente (Profesores, alumnos de la diplomatura, licenciatura y tercer ciclo y Personal de Administración y Servicios (PAS)). Los encuestados fueron elegidos al azar en diversos momentos del día, mientras hacían uso de la red.
2. Selección de buscadores. Han sido seleccionados cuatro de los cuales dos son motores de búsqueda de primera generación (Alta Vista y FAST Search), uno de segunda generación (Northern Light) y un directorio (Yahoo) en el cual las búsquedas se han realizado a través de browsing. Esta selección responde a intereses de los evaluadores debido al uso que se hace ellos en su trabajo diario.
3. Test de evaluación. Se ha realizado una selección de todas las preguntas recogidas y elección de las expresiones exactas a los que serán sometidos los agentes buscadores.

4. Método de búsqueda. Se han realizado dos tipos de búsqueda, por lenguaje natural y con operadores booleanos, según lo que permita el buscador correspondiente.
5. Método de evaluación. Se analizarán las posibilidades de búsqueda a partir de un cuadro de evaluación que considere las siguientes características:
 - Evaluación de las características de la pantalla de búsqueda.
 - Evaluación de las características de la pantalla de resultados: búsquedas en lenguaje natural, elegir el número de resultados por página, elegir el nivel de descripción de los resultados, ayudas derivadas de los resultados, descender a través de las categorías, controlar el grado de relevancia en los términos, elegir el idioma de las páginas, limitar por fechas, entre otros.
 - Evaluación de las posibilidades de consulta con operadores booleanos: truncamientos (*), frases exactas (".."), AND ("+"), OR, NOT("-") y NEAR, entre otros.
6. En el momento de realizar las búsquedas se ha elaborado un cuadro en el que se analizan cada una de las preguntas en todos los buscadores y desde el punto de vista de la búsqueda en texto libre / lenguaje natural y con los operadores booleanos, con la finalidad de demostrar como se obtiene la mayor precisión en las búsquedas y cuales son los resultados que se obtienen. Siempre se ha empleado la opción de búsqueda avanzada en todos los buscadores y los resultados de la **Exhaustividad, Relevancia y Pertinencia** se obtienen a partir de los operadores booleanos.
7. Para finalizar se ha elaborado un cuadro-resumen que trata de mostrar conclusiones generales de cada buscador en función de todas las variables analizadas. En este cuadro se especifica si el buscador ha dado resultados y se evalúa con una puntuación de 1-3, considerándolo de menor importancia (1) a mayor (3).

3. Descripción y características de los buscadores

3.1. ALTA VISTA

AltaVista fue desarrollado por Digital Equipment Corporation en 1995 en laboratorios de investigación de Palo Alto, es un motor de búsqueda para buscar todos los documentos publicados en la WWW. Usando AltaVista, se puede encontrar cualquier palabra en cualquier documento publicado en el World Wide Web o en grupos de discusión del USENET, que hace que AltaVista sea una herramienta inestimable para encontrar información. Actualmente tiene indizadas 200 millones de páginas.

Su característica fundamental es la entrega de plataformas múltiples de búsqueda, programación en tiempo real, compras y portal de servicios locales.

Otras características del motor son las primeras búsquedas multilingües en Internet, que se hicieron desde su creación. Alta Vista desarrolla también el Babel Fish, el primer servicio de Internet del Web que traduce palabras, frases, y sitios enteros del Web en línea, en diversos idiomas como el español, francés, alemán, portugués e italiano. Recientemente, lanzó un buscador de fotografías, tecnología de búsqueda de la imagen, y el filtro que reduce el ruido de los resultados de las búsquedas. Otras mejoras incluyen la detección de la frase, revisión de palabras del lenguaje natural. También permite capacidades de búsqueda de elementos multimedia.

POSIBILIDADES	ALTA VISTA
<i>Búsquedas simples</i>	Sí
<i>Búsquedas avanzadas</i>	Sí
<i>Búsquedas en lenguaje natural</i>	Reconoce palabras con significado. Suprime palabras sin significado (artículos, preposiciones, adverbios,...)
<i>Elegir el nº de resultados por página</i>	No. Sólo muestra 10 resultados por página hasta 20 páginas.
<i>Ejemplos de búsquedas</i>	Sí, antes de formular la búsqueda
<i>Descripción de los buscadores</i>	Sí, más detallada en búsqueda avanzada.
<i>Sistemas de ayuda</i>	Sí
<i>Autor y Contactos</i>	Sí
<i>Diseño de la página (directorios, noticias, servicios adicionales, publicidad, banners)</i>	Directorios Servicios adicionales Publicidad Banner en la parte superior

TABLA 1. Características de la pantalla de búsqueda.

POSIBILIDADES	ALTA VISTA
<i>Ranking u orden de los resultados</i>	En la búsqueda avanzada, permite especificar fechas, idioma, y muestra en primer lugar frases o palabras que se indiquen en la opción "Sort by". No controla el grado de relevancia.
<i>Ayudas derivadas de los resultados (descripción, resumen, ...)</i>	Ofrece términos de una lista de palabras clave. Muestra el título y el primer párrafo de la página, el URL y la posibilidad de traducción ("Translate"). Ofrece términos relacionados y otras páginas que pertenecen al mismo URL.
<i>Retroalimentación (feedback)</i>	Sí, "Related pages", al final de la página sugiere fuentes alternativas de búsqueda.
<i>Descender a través de categorías</i>	No
<i>Elegir el idioma de las páginas</i>	Da posibilidad de elegir entre 25 idiomas.
<i>Limitar por fechas</i>	Sí por DD/MM/AA
<i>Búsqueda por imágenes, audio y video</i>	En la pantalla principal existe la posibilidad de seleccionar las búsquedas por IMAGES, MP3/AUDIO, VIDEO.
<i>Búsqueda por campos</i>	Permite buscar por TITLE: "frase"; TEXT: "frase"; ANCHOR: "frase" (busca enlaces con nombre propio, por ejemplo: ANCHOR: "tesauro de educación"); DOMAIN: ".dominio"; IMAGE: "nombre" o "nombre.extensión"; APPLE: "nombre" (para búsquedas que contienen aplicaciones JAVA)
<i>Búsqueda por HOST</i>	Permite buscar por HOST: "frase"
<i>Búsqueda por URL</i>	Permite buscar por URL: "frase"
<i>Búsqueda por LINK</i>	Permite buscar por LINK: "nombre.extensión".

TABLA 2. Características de la pantalla de resultados de la búsqueda.

POSIBILIDADES	ALTA VISTA
<i>Truncamiento (*)</i>	Permite truncamientos en el medio y al final de la palabra y reconoce como mínimo tres caracteres. Puede reemplazar de 0 a 5 caracteres.
<i>Frases exactas ("..")</i>	Utiliza las comillas.
<i>AND</i>	Utiliza el operador AND y el signo "+".
<i>OR</i>	Utiliza el operador OR.
<i>NOT</i>	Utiliza el operador NOT o el signo "-".
<i>NEAR</i>	Utiliza el operador NEAR o "~", aproximando 10 palabras en búsqueda avanzada y 8 en búsqueda simple.
<i>Reconoce signos lingüísticos (sensibles e insensibles)</i>	Es insensible, no considera: comas (","), puntos ("."); guión ("-"); subrayado ("_"); barras ("/"), acentos, minúsculas. Es sensible a las palabras escritas con mayúscula o con palabras que utilizan ambas.

TABLA 3. Posibilidades de consulta con Operadores Booleanos.

3.2. FAST SEARCH

Fast Search & Transfer ASA (FAST) se estableció el 16 de julio de 1997 en Noruega y abrió una subsidiaria americana en 1998. La investigación y desarrollo es realizada en Oslo y Trondheim en Noruega.

FAST tiene como meta de su tecnología de búsqueda, construir un motor de búsqueda más grande del mundo y por eso la frase que la caracteriza "All the web, All the Time" (toda la red, todo el tiempo). Actualmente FAST clasifica 575 millones de sitios y posee servidores Dell Power Edge y subsistemas de almacenamiento Power Vaultt, estos servidores son ejecutados con el sistema operativo BSD.

POSIBILIDADES	FAST SEARCH
Búsquedas simples	Sí
<i>Búsquedas avanzadas</i>	Sí
<i>Búsquedas en lenguaje natural</i>	Sí. Busca en primer lugar la frase y palabras sueltas sino se ponen las comillas
<i>Elegir el nº de resultados por página</i>	Sí, en la búsqueda avanzada. En la búsqueda simple aparecen de 10 en 10 y en la búsqueda avanzada hasta 100 resultados.
<i>Ejemplos de búsquedas</i>	No.
<i>Descripción de los buscadores</i>	Sí, pero muy simple.
<i>Sistemas de ayuda</i>	No.
<i>Autor y Contactos</i>	Sí
<i>Diseño de la página (directorios, noticias, servicios adicionales, publicidad, banners)</i>	-

TABLA 1. Características de la pantalla de búsqueda.

POSIBILIDADES	FAST SEARCH
<i>Ranking u orden de los resultados</i>	Sí lo hace en la búsqueda avanzada priorizando una frase o palabra, carga o descarga una frase o palabra y da hasta cuatro posibilidades. Además, estas frases o palabras pueden buscarse en el texto, título, URL, LINK o en ambas. Excluye o incluye dominios (.es, .com, ...) No controla el grado de relevancia.
<i>Ayudas derivadas de los resultados (descripción, resumen, ...)</i>	Presenta el título y el primer párrafo de la página y el URL.
<i>Retroalimentación (feedback)</i>	No.
<i>Descender a través de categorías</i>	No.
<i>Elegir el idioma de las páginas</i>	Da la posibilidad de elegir entre 31 idiomas en búsqueda avanzada.
<i>Limitar por fechas</i>	No
<i>Búsqueda por imágenes, audio y video</i>	Da la posibilidad de buscar por PICTURES SEARCH o MULTIMEDIA SEARCH (imágenes, vídeos, streams, audio); MP3/SEARCH; FTP SEARCH (busca archivos de sonido: .av, .mp3, .midi; pictures: .jpg, .gif).
<i>Búsqueda por campos</i>	Sí, dentro de la búsqueda avanzada por TEXT, TITLE (marcando el recuadro correspondiente)
<i>Búsqueda por HOST</i>	No
<i>Búsqueda por URL</i>	Sí, dentro de la búsqueda avanzada (marcando el recuadro correspondiente)
<i>Búsqueda por LINK</i>	Sí, dentro de la búsqueda avanzada (marcando el recuadro correspondiente)

TABLA 2. Posibilidades de la pantalla de resultados.

POSIBILIDADES	FAST SEARCH
<i>Truncamiento (*)</i>	No
<i>Frases exactas ("..")</i>	Utiliza las comillas. También se puede seleccionar como una forma de búsqueda dentro de la opción "The Exact Phrase".
<i>AND</i>	Utiliza el signo "+".
<i>OR</i>	No
<i>NOT</i>	Utiliza el signo "-".
<i>NEAR</i>	No
<i>Reconoce signos lingüísticos (sensibles e insensibles)</i>	Es sensible a signos lingüísticos como el acento. Es insensible a las mayúsculas, minúsculas, comas (","), guiones ("-"), subrayados ("_"), &.

TABLA 3. Posibilidades de consulta con Operadores Booleanos.

3.3. NORTHERN LIGHT

Northern Light Technology, Incorporated fue premiado como el primer motor de búsqueda en agosto de 1997 y desde ese momento ha crecido considerablemente. Se trata de un motor de búsqueda de segunda generación que permite buscar información en la World Wide Web en más de 5400 fuentes distintas. Los resultados de la Web se combinan con información sacada de otras fuentes dando acceso a monografías, revistas, bases de datos y otras fuentes que no son accesibles desde otros motores de búsqueda (es lo que se denomina como "Internet Invisible"). Además los resultados de las búsquedas aparecen distribuidos en carpetas que facilitan la búsqueda. En la actualidad tiene indizadas 220 millones de páginas.

POSIBILIDADES	NORTHERN LIGHT
<i>Búsquedas simples</i>	Sí
<i>Búsquedas avanzadas</i>	Sí
<i>Búsquedas en lenguaje natural</i>	Sí, permite la búsqueda haciendo una pregunta en lenguaje natural o con varias palabras simples.
<i>Elegir el nº de resultados por página</i>	No.
<i>Ejemplos de búsquedas</i>	Sí, en las ayudas del motor ("tips")
<i>Descripción de los buscadores</i>	Sí, en la página principal ("About")
<i>Sistemas de ayuda</i>	Sí, no da mucha información.
<i>Autor y Contactos</i>	Sí
<i>Diseño de la página (directorios, noticias, servicios adicionales, publicidad, banners)</i>	No tiene directorio para hacer búsquedas por browsing. Se pueden consultar noticias y publicidad sobre asuntos de actualidad o información sobre el propio programa, últimas novedades,... No tiene banners.

TABLA 1. Características de la pantalla de búsqueda.

POSIBILIDADES	NORTHERN LIGHT
<i>Ranking u orden de los resultados</i>	En la búsqueda avanzada se puede seleccionar la materia, el tipo de documento, idioma, país y fecha. Controla la relevancia automáticamente.
<i>Ayudas derivadas de los resultados (descripción, resumen, ...)</i>	Título de la página; porcentaje de relevancia; tipo de documento según el buscador; breve resumen; fecha de indización; tipo de recurso web que existe; dirección del site; carpeta que agrupa todos los sites derivados de esa página principal.
<i>Retroalimentación (feedback)</i>	Sí. A parte de los resultados principales, agrupa la información en carpetas bajo otras materias que son de interés.
<i>Descender a través de categorías</i>	No. Agrupa los resultados en carpetas por temas según la búsqueda realizada.
<i>Elegir el idioma de las páginas</i>	Sí, en la búsqueda avanzada permite hasta 5 idiomas.
<i>Limitar por fechas</i>	Sí, por MM/DD/AA
<i>Búsqueda por imágenes, audio y video</i>	No.
<i>Búsqueda por campos</i>	Sí, permite la búsqueda por: TITLE: frase; PUB: título de la revista; COMPANY: compañía particular; TICKER: company's stock ticker; TEXT: frase; RECID: Special Collection document; SORT: date
<i>Búsqueda por HOST</i>	No
<i>Búsqueda por URL</i>	Sí, URL: nombre.extensión
<i>Búsqueda por LINK</i>	No

TABLA 2. Características de la pantalla de resultados.

POSIBILIDADES	NORTHERN LIGHT
<i>Truncamiento (*)</i>	Sí, emplea los signos "*" y "%".
<i>Frases exactas ("..")</i>	Sí, emplea el signo "...".
<i>AND</i>	Sí, emplea el operador AND y el signo "+".
<i>OR</i>	Sí, emplea el operador OR
<i>NOT</i>	Sí, emplea el operador NOT y el signo "-".
<i>NEAR</i>	Sí, utiliza el signo "...".
<i>Reconoce signos lingüísticos (sensibles e insensibles)</i>	Es insensible a signos ortográficos como acentos ("^"), guiones ("-"), subrayados ("_"), comas (","), puntos ("."), mayúsculas, minúsculas,...

TABLA 3. Posibilidades de consulta con Operadores Booleanos.

3.4. YAHOO

Yahoo! significa "Yat Another Hierarchical Officious Oracle", fue creada por David Filo y Jerry Yang, estudiantes de doctorado de Ingeniería Eléctrica de la Universidad de Stanford, comenzaron con la guía de Yahoo en 1994, como un modo de hacer un seguimiento de intereses personales y luego las listas se hicieron difíciles de manejar, crearon una base de datos y crearon un software personalizado para identificar y editar de forma eficaz el material almacenado en Internet.

Tiene una cobertura internacional con nodos locales que sirven a sus respectivos ámbitos geográficos que incluye a ocho países europeos y a otros continentes (por ejemplo la versión para Japón). Para Ibero América existen una versión española, una mexicana y una brasileña.

Yahoo no es un motor de búsqueda. Si no que es una importante fuente de información de Web. Trabaja como un indexado jerárquico de temas, permitiendo la penetración desde lo general a lo específico. Yahoo es un intento de organizar y catalogar la Web.

La recogida de páginas o sites son sugeridos por los propios usuarios, por lo tanto no es un directorio muy selectivo, mas tarde son visitados y evaluados

por Yahoo, existe un formulario de sugerencias donde se solicita la URL, el título y descripción del recurso a incluir, muchos sitios son proporcionados por las mismas compañías, debido a que Yahoo es muy conocido en la Web, su cobertura de sitios comerciales es especialmente buena. Yahoo esta organizado por categorías temáticas e incluye selecciones de la semana, novedades, noticias y el famoso “cool” seleccionado de las categorías específicas dentro de Yahoo, que tiene un contenido divertido, extraordinario y especialmente útil.

Yahoo dispone de un sistema de puntuación, los sitios que disponen de una etiqueta en la que aparece la imagen de las gafas son las que se consideran de un contenido de calidad. Y deben realizarse en primer orden. La etiqueta ¡NEW! indica que la entrada fue añadida la semana anterior, si aparece junto a la entrada de un directorio, significa que en las entradas de la jerarquía de ese directorio se han modificado en los últimos tres días. Cuando la etiqueta “@” aparece al final de una categoría significa que ese titular está mencionado en otros lugares dentro de la jerarquía de Yahoo!, si se selecciona ese titular, nos llevará a su ubicación principal. Los números que aparecen entre paréntesis junto a las categorías de Yahoo indican el número de entradas clasificadas bajo ese encabezamiento de entrada.

Actualmente tiene indizadas unos 60 millones de páginas.

POSIBILIDADES	YAHOO
Búsquedas simples	Sí
Búsquedas avanzadas	Sí
Búsquedas en lenguaje natural	Sí, dentro de la búsqueda avanzada
Elegir el nº de resultados por página	Sí, en la búsqueda avanzada. Permite que después de la primera página se puedan presentar hasta 100 resultados por página teniendo como mínimo 20.
Ejemplos de búsquedas	Sólo en la búsqueda avanzada en al ayuda del directorio
Descripción de los buscadores	Sí, en la página principal.
Sistemas de ayuda	Sí
Autor y Contactos	Sí
Diseño de la página (directorios, noticias, servicios adicionales, publicidad, banners)	Es un directorio aunque ofrece posibilidades de búsqueda por palabras clave. El directorio tiene 14 categorías interactivas que permites ir navegando por la información. Dentro de cada categoría existe la posibilidad de una búsqueda por palabras clave. Tiene información sobre noticias de actualidad y sobre el propio directorio (guías de Yahoo, directorios de lugares del mundo, ayudas de búsqueda, información sobre los fabricantes,...) Ofrece servicios adicionales como deportes, televisión, tiempo, información sobre negocios, trabajo, coches, e-mail, chat,... En los banners aparece la publicidad.

TABLA 1. Características de la pantalla de búsqueda.

POSIBILIDADES	YAHOO
<i>Ranking u orden de los resultados</i>	No.
<i>Ayudas derivadas de los resultados (descripción, resumen, ...)</i>	Aparece el descenso de categorías; título de la página; breve resumen del contenido
<i>Retroalimentación (feedback)</i>	No.
<i>Descender a través de categorías</i>	Sí. Ofrece la posibilidad de hacer "browsing" a través del directorio. Combina ambas posibilidades ya que se puede hacer un browsing y luego limitar la búsqueda por palabras clave o al contrario. Los resultados de la búsqueda por palabras clave se organizan por categorías.
<i>Elegir el idioma de las páginas</i>	No
<i>Limitar por fechas</i>	Sí, permite búsquedas con tres años de antigüedad como máximo y de un día como mínimo.
<i>Búsqueda por imágenes, audio y video</i>	No
<i>Búsqueda por campos</i>	Sí, búsqueda por title: frase o t: frase
<i>Búsqueda por HOST</i>	No
<i>Búsqueda por URL</i>	No
<i>Búsqueda por LINK</i>	No

TABLA 2. Características de la pantalla de resultados.

POSIBILIDADES	YAHOO
<i>Truncamiento (*)</i>	Utiliza el signo "***"
<i>Frases exactas ("..")</i>	Utiliza el signo "..."
<i>AND</i>	Sí, es una opción dentro de la búsqueda avanzada. Emplea el signo "+".
<i>OR</i>	Sí, es una opción dentro de la búsqueda avanzada.
<i>NOT</i>	Utiliza el signo "-".
<i>NEAR</i>	No.
<i>Reconoce signos lingüísticos (sensibles e insensibles)</i>	Es insensible a signos ortográficos como acentos ("´"), guiones ("`"), subrayados ("_"), comas (","), puntos ("."), mayúsculas, minúsculas,...

TABLA 3. Posibilidades de consulta con Operadores Booleanos.

4. Elaboración de los enunciados o sintaxis de búsqueda

De todas las preguntas que se recopilaron entre la comunidad universitaria se han seleccionado cinco. Estas son:

<i>PREGUNTAS SOBRE OCIO</i>	<i>PREGUNTAS PARA ACTIVIDADES PROFESIONALES</i>	<i>PREGUNTAS CIENTÍFICAS</i>
Turismo (Viajes Salamanca,...)	Información sobre Atapuerca	Revistas electrónicas sobre biblioteconomía I+D público en Europa Libertad intelectual, Derecho a la información y Libertad de expresión en bibliotecas

Posteriormente se procedió a desarrollar la sintaxis de búsqueda de cada una de las preguntas tanto en texto libre o lenguaje natural como con operadores booleanos.

4.1. Texto libre / lenguaje natural

1. turismo en Salamanca, Salamanca tourist, tourist in Salamanca
2. Atapuerca yacimientos arqueológicos, Atapuerca fossils
3. revista* electrónica* sobre bibliotec*, revista* electrónica* en bibliotec*, electronic journal* on librar*, electronic journal* in librar*
4. I+D público en Europa, investigación y desarrollo en Europa, R&D public in Europe, research and development public in Europe
5. libertad intelectual en bibliotec*, derecho a la información en bibliotec*, intellectual freedom in librar*, law of information in librar*

4.2. Operadores booleanos

1. (turismo OR tourist) AND Salamanca
2. Atapuerca AND (yacimientos arqueológicos OR fossils)
3. ((revista* OR journal*) AND (electrónica* OR electronic*)) AND (bibliotec* OR librar*)
4. (("I+D" OR "Investigación y Desarrollo") AND public* AND Europ*) OR (("R&D" OR "research and development") AND public* AND Europ*)
5. ("libertad intelectual" OR "derecho de información" OR "intellectual freedom" OR "law of information") AND (bibliotec* OR librar*)

5. Evaluación de los resultados de las búsquedas

El cuadro que aparece al final de este apartado contiene los resultados que se obtienen de las preguntas en lenguaje natural y con operadores booleanos, así

como las fechas en que fueron realizadas, organizados por buscadores. Los cálculos de exhaustividad, relevancia y pertinencia están hechos a partir de las búsquedas con operadores booleanos. Para poder calcular la **Exhaustividad** se tiene en cuenta el número total de páginas que indiza cada buscador y se hace un cálculo aproximado para cada uno de ellos según cálculo de Notes¹; Alta Vista 200 millones, Fast Search 300 millones, Northern Light 220 millones y Yahoo 60 millones. Para la **Relevancia** se analizan los 20 primeros resultados de cada búsqueda tomando en consideración los duplicados e inactivos. En el caso de la **Pertinencia** se tiene en cuenta el contenido de cada uno de los resultados. Se han tenido en cuenta tres posibilidades:

- direcciones que tratan el tema puntualmente (sólo estos datos aparecen en el cuadro),
- direcciones que tratan el tema juntamente con otros,
- direcciones que para nada lo tratan.

Para cada uno de los resultados se ha calculado el porcentaje respectivo para establecer más claramente la eficiencia de los buscadores analizados.

En la búsqueda en lenguaje natural o texto libre algunas veces aparecen dos resultados debido a que la búsqueda se ha realizado de dos formas diferentes (Véase en apartado 3. Sintaxis de búsqueda), se han omitido las búsquedas en inglés, considerando que los resultados de las búsquedas en español son suficientemente significativas.

5.1. Análisis de las preguntas propuestas

Analizando los resultados de todas las preguntas podemos afirmar que FAST SEARCH es el buscador que mayor número de páginas muestra en lenguaje natural (4 de 5 preguntas), esto se debe a que el motor lee la frase completa o parte/ s de ella alternativamente, seguido por NORTHERN LIGHT que muestra mayor número de registros en una sola pregunta. Estas variables cambian cuando se realizan las búsquedas con operadores booleanos donde ALTA VISTA es el que mayor número de registros recupera (2 de 5 preguntas), seguido de los otros buscadores con una pregunta cada uno.

Las diferencias entre las búsquedas por lenguaje natural y operadores booleanos muestran diferencias significativas. Mientras que en algunas preguntas las búsquedas en lenguaje natural representan un porcentaje inferior a los obtenidos con los operadores booleanos, en otras es muy superior. Esto

¹ Greg R. Notes. "Search Engine Statistics: Database Total Size Estimates". [En línea]. <http://www.notess.com/search/stats/sizeest.shtml>. [Consultado el 23 de mayo de 2000].

viene a corroborar las características de cada uno de los buscadores, ya que en el caso de FAST SEARCH a pesar de tener más desarrollada la opción de búsqueda por texto libre, muestra resultados muy amplios en dos preguntas (Libertad Intelectual e I+D público) en el que muestra cifras entre 21 y 183 millones para la primera pregunta y 26 y 32 millones para la segunda. Comparándolo con los resultados en la búsqueda con operadores booleanos son sólo el 0,05% y 0,005% para el primer caso y 2,4% y 1,9 % para el segundo. El caso contrario a esto lo demuestra ALTA VISTA que en las búsquedas por texto libre sobre Revistas electrónicas, Atapuerca y Libertad intelectual muestran porcentajes muy pequeños, 0% para la primera pregunta, 0,4% para la segunda y 2,38 % para la tercera.

En cuanto a la **relevancia** de los buscadores los que mayor número de registros inactivos muestra son ALTA VISTA (11) y NORTHERN LIGHT (9), seguidos de YAHOO y FAST SEARCH con 7. El buscador que más duplicados muestra es FAST SEARCH (40) seguidos de NORTHERN LIGHT (4) y ALTA VISTA (3). YAHOO no tiene registros duplicados debido a que sus mejores posibilidades de búsqueda se realizan por browsing y por lo tanto los resultados se ofrecen agrupados por categorías.

La **pertinencia** de los primeros 20 resultados muestra que NORTHERN LIGHT y YAHOO obtienen mejores resultados en todas las preguntas, aunque este último ha realizado sus búsquedas a través de browsing. Es necesario resaltar que FAST SEARCH obtiene en algunos casos resultados relevantes (Libertad intelectual 100%). Teniendo en cuenta los resultados de registros que tratan el tema juntamente con otros, se puede ver que YAHOO y NORTHERN LIGHT obtienen porcentajes que fluctúan entre el 60% (Libertad intelectual) y 30% (Revistas electrónicas e I+D público) respectivamente. El buscador que muestra la peor pertinencia es ALTA VISTA que en el mejor de los casos obtiene un 20% (Libertad intelectual) y en otras tres preguntas sólo obtiene el 0% de pertinencia.

A medida que se han ido realizando las búsquedas se comprobó que si se emplean operadores booleanos con “preguntas científicas” los resultados no son pertinentes, por lo que todos los casos es necesario fraccionar las búsquedas de varias formas para obtener una mayor pertinencia. Para mejorar las búsquedas se pueden emplear las distintas opciones de búsqueda avanzada como “Sort by” en Alta Vista, “Must include” en Fast Search y “Subject”, “Type of Document” y “Sort” en Northern Light.

HERRAMIENTAS DE BÚSQUEDA	PREGUNTAS	FECHA DE BÚSQUEDA	TEXTO LIBRE/LENGUAJE NATURAL	OPERADORES BOOLEANOS			
				Nº DE RESULTADOS (EXHAUSTIVIDAD)	RELEVANCIA (20 PRIMEROS RESULTADOS)		PERTINENCIA (20 PRIMEROS RESULTADOS)
					DUPLICADOS	INACTIVOS	
ALTA VISTA	Turismo en Salamanca	19-05-2000	13	8250 (0,01%)	1 (5%)	2 (10%)	0 (0%)
	Atapuerca	20-05-2000	1	239 (0,01%)	1 (5%)	4 (20%)	1 (5%)
	Revistas electrónicas en Biblioteconomía	20-05-2000	22 / 9	921124 (0,46%)	0 (0%)	3 (15%)	0 (0%)
	Libertad intelectual, Derecho a la información en bibliotecas	23-05-2000	0 / 0	16136 (0,01%)	0 (0%)	1 (5%)	4 (20%)
	I+D público en Europa	20-05-2000	0 / 0	1.516.127 (0,76%)	1 (5%)	2 (10%)	0 (0%)
FAST SEARCH	Turismo en Salamanca	19-05-2000	7050	9649 (0,003%)	9 (45%)	4 (20%)	10 (50%)
	Atapuerca	20-05-2000	132	302 (0,0001%)	8 (40%)	1 (5%)	14 (70%)
	Revistas electrónicas en Biblioteconomía	20-05-2000	2 / 3	626 (0,0002%)	4 (20%)	1 (5%)	1 (5%)
	Libertad intelectual, Derecho a la información en bibliotecas	23-05-200	21.320.038 / 183.221.951	9786 (0,003%)	5 (25%)	0 (0%)	20 (100%)
	I+D público en Europa	20-05-2000	26.766.689 / 32.449.203	635998 (0,21%)	14 (70%)	1 (5%)	0 (0%)

HERRAMIENTAS DE BÚSQUEDA	PREGUNTAS	FECHA DE BÚSQUEDA	TEXTO LIBRE/LENGUAJE NATURAL	OPERADORES BOOLEANOS			
				Nº DE RESULTADOS (EXHAUSTIVIDAD)	RELEVANCIA (20 PRIMEROS RESULTADOS)		PERTINENCIA (20 PRIMEROS RESULTADOS)
					DUPLICADOS	INACTIVOS	
NORTHERN LIGHT	Turismo en Salamanca	19-05-2000	3152	4989 (0,01%)	0 (0%)	1 (5%)	17 (85%)
	Atapuerca	20-05-2000	114	350 (0,0001%)	2 (10%)	3 (15%)	13 (65%)
	Revistas electrónicas en Biblioteconomía	20-05-2000	8184 / 9671	88526 (0,04%)	0 (0%)	2 (10%)	4 (20%)
	Libertad intelectual, Derecho a la información en bibliotecas	23-05-2000	6867 / 241149	46644 (0,02%)	2 (10%)	2 (10%)	10 (50%)
	I+D público en Europa	23-05-2000	11199 / 24319	192478 (0,09%)	0 (0%)	1 (5%)	4 (20%)
YAHOO	Turismo en Salamanca	23-05-2000	1	16 (0,00002%)	0 (0%)	4 (20%)	4 (20%)
	Atapuerca	23-05-2000	18	614 (0,001%)	0 (0%)	1 (5%)	18 (90%)
	Revistas electrónicas en Biblioteconomía	23-05-2000	16	17 (0,00003%)	0 (0%)	0 (0%)	20 (100%)
	Libertad intelectual, Derecho a la información en bibliotecas	23-05-2000	326 / 7	41 (0,00007%)	0 (0%)	0 (0%)	7 (35%)
	I+D público en Europa	23-05-2000	0 / 2	5 (0,000008%) / 88 (0,0001%)	0 (0%)	2 (10%)	6 (30%)

6. Conclusiones

Se ha elaborado una tabla final con ponderaciones de cada uno de los buscadores en función de las características analizadas. Las ponderaciones se han considerado con una puntuación del 1 al 3, siendo el 1 la más baja y el 3 la más alta. En aquellas características que los buscadores no desarrollan se ha optado por no dar ningún tipo de puntuación.

Alta Vista es un motor de primera generación que tiene mucha versatilidad ya que en la evaluación de la pantalla de búsqueda y de la página de resultados cumple satisfactoriamente ocho de diez características en el primer caso y todas en el segundo. Trabaja con muchas combinaciones de operadores booleanos, pero los resultados de las búsquedas no son satisfactorios (baja pertinencia) a pesar de tener una alta exhaustividad.

Fast Search es un motor de primera generación que se caracteriza por una presentación sencilla como lo demuestra la tabla, aunque tiene sistemas de ayuda que están poco desarrollados. En cuanto a las búsquedas con operadores booleanos no permite realizar muchas combinaciones, y los resultados que ofrece no son agrupados por URL, lo cual hace que presente muchos duplicados. Sin embargo su nivel de exhaustividad y pertinencia es alto.

Northern Light es un motor de segunda generación que agrupa los resultados en función del porcentaje de relevancia del registro recuperado. Ofrece posibilidades de retroalimentación o feedback ya que los resultados son agrupados en carpetas según cada posible temática de los términos de la búsqueda. Combinan bien los operadores booleanos y se tiene una alta pertinencia.

Yahoo es un directorio que ofrece mejores resultados a través de la combinación de browsing y búsquedas, por lo tanto tiene un bajo rendimiento en búsquedas simples y avanzadas. No soporta muchas combinaciones de operadores booleanos, pero los resultados por browsing son óptimos teniendo en cuenta su nivel de pertinencia.

Cada buscador tiene potencialidades diversas para opciones diferentes y cuando se desea resolver preguntas complejas dificulta la obtención de resultados óptimos para el usuario común ya que para ello se necesita tener un conocimiento profundo del buscador en el que se va a realizar la búsqueda.

<i>CARACTERÍSTICAS</i>		<i>ALTA VISTA</i>	<i>FAST SEARCH</i>	<i>NORTHERN LIGHT</i>	<i>YAHOO</i>
CARACTERÍSTICAS DE LA PANTALLA DE BÚSQUEDA	Búsquedas simples	✓ 2	✓ 3	✓ 3	✓ 2
	Búsquedas avanzadas	✓ 3	✓ 2	✓ 3	✓ 1
	Búsquedas en lenguaje natural	✓ 1	✓ 3	✓ 3	✓ 1
	Elegir el nº de resultados por página	•	✓ 3	•	✓ 3
	Ejemplos de búsquedas	✓ 3		✓ 2	✓ 1
	Descripción de los buscadores	✓ 3	✓ 1	✓ 2	✓ 1
	Sistemas de ayuda	✓ 3	✓ 1	✓ 2	✓ 1
	Autor y Contactos	✓ 3	✓ 3	✓ 3	✓ 3
	Diseño de la página (directorios, noticias, servicios adicionales, publicidad, banners)	✓ 2	✓ 3	✓ 2	✓ 2
	CARACTERÍSTICAS DE LA PÁGINA DE RESULTADOS	Ranking u orden de los resultados	✓ 2	✓ 1	✓ 2
Ayudas derivadas de los resultados (descripción, resumen, ...)		✓ 2	✓ 1	✓ 3	•
Retroalimentación (feedback)		✓ 2	•	✓ 3	•
Descender a través de categorías		✓ 1	•	•	✓ 3
Elegir el idioma de las páginas		✓ 2	✓ 3	✓ 1	•
Limitar por fechas		✓ 3		✓ 3	✓ 2
Búsqueda por imágenes, audio y vídeo		✓ 3	✓ 3	•	•
Búsqueda por campos		✓ 3	✓ 1	✓ 3	✓ 1
Búsqueda por HOST		✓ 3	•	•	•
Búsqueda por URL		✓ 3	✓ 3	✓ 3	•
Búsqueda por LINK		✓ 3	✓ 3	•	•

CARACTERÍSTICAS		ALTA VISTA	FAST SEARCH	NORTHERN LIGHT	YAHOO
OPERADORES BOOLEANOS	Truncamiento (*)	✓ 3	•	✓ 2	✓ 1
	Frases exactas ("...")	✓ 3	✓ 3	✓ 2	✓ 2
	AND	✓ 3	✓ 3	✓ 3	✓ 2
	OR	✓ 3	✓ 1	✓ 2	✓ 1
	NOT	✓ 3	✓ 3	✓ 3	✓ 3
	NEAR	✓ 3	•	✓ 3	•
	Reconoce signos lingüísticos (sensibles e insensibles)	✓ 3	✓ 2	✓ 3	✓ 3
RESPUESTAS	Exhaustividad	✓ 3	✓ 2	✓ 2	✓ 2
	Relevancia	✓ 2	✓ 1	✓ 3	✓ 3
	Pertinencia	✓ 1	✓ 3	✓ 3	✓ 3

BIBLIOGRAFÍA

- **HÍPOLA**, Pedro; **VARGAS QUESADA**, Benjamín; **MONTES HERNÁNDEZ**, Agustín. “Descripción y evaluación de agentes multibuscadores”. *El Profesional de la Información*. 8/11 (1999). P. p. 15-26.
- **MALDONADO MARTÍNEZ**, Elena; **FERNÁNDEZ SÁNCHEZ**, Elena. “Evaluación de los principales “buscadores” desde un punto de vista documental: recogida, análisis y recuperación de los recursos de información. En: *Actas VI Jornadas de Documentación*. Madrid: FESABID, 199?.—p. 529-548.
- **MONTES HERNÁNDEZ**, Agustín. “Posibilidades de consulta en los buscadores”. *El Profesional de la Información*. 8/3 (1999). p. p. 8-14.
- **NOTES**, Greg R. “Search Engine Statistics: Database Total Size Estimates”. [En línea]. <http://www.notess.com/search/stats/sizeest.shtml>. Consultado el 23 de mayo de 2000.
- **OLVERA LOBO**, M^a Dolores. “Rendimiento de los sistemas de recuperación de información en la World Wide Web: Revisión metodológica”. *Revista Española de Documentación Científica*. 23/1 (2000). p.p. 63-77.
- **ROUSSEAU, Ronald**. “Daily time series of common single word searches in Alta Vista and NorthernLight” [En línea]. *Cybermetrics*. 2/3 (1998). <http://www.cindoc.csic.es/cybermetrics/cybermetrics/>.
- **SENSO**, José A. “Herramientas para realizar búsquedas en Internet: una revisión”. *El Profesional de la Información*. 7/1-2 (1998). P..p. 24-25.

SOBRE LOS AUTORES

Martín Alonso Estrada Cuzcano (Lima, 1964) Licenciado en Bibliotecología de la UNMSM, y estudiante del Doctorado en Documentación en la Universidad Carlos III de Madrid. Ha publicado diversos artículos en publicaciones especializadas, y también ha participado como ponente en diversos eventos nacionales e internacionales. Actualmente trabaja en la Biblioteca Central de la Universidad de Piura.

Noelia Rodríguez Maniega (León, 1977) Licenciada en Documentación por la Universidad Carlos III de Madrid y realiza estudios de Doctorado en la misma universidad. Actualmente trabaja como Documentalista en la empresa SOFRES: Documentación y Análisis.