

Enl@ce: Revista Venezolana de Información,
Tecnología y Conocimiento
ISSN: 1690-7515
Depósito legal pp 200402ZU1624
E-ISSN: 2542-3274
Depósito legal ppi 201502ZU4693
Año 14: No. 1, Enero-Abril 2017, pp. 29-50

Cómo citar el artículo (Normas APA):

Amaya, M. (2017). Evaluación de herramientas de recuperación de información electrónica. *Enl@ce Revista Venezolana de Información, Tecnología y Conocimiento*, 14 (1), 29-50

Evaluación de herramientas de recuperación de información electrónica

Evaluation of Electronic Information Retrieval Tools

*Miguel Ángel Amaya Ramírez*¹

Resumen

El crecimiento exponencial de los volúmenes de información disponibles a través del internet, conlleva a que los usuarios de la red de redes, utilicen las herramientas de búsqueda, situación que conlleva a enfrentar varios problemas al momento de recuperar los contenidos requeridos. En función de ello, se procede a evaluar el funcionamiento de diferentes herramientas de búsqueda y recuperación de la información que se encuentran en la web. Mediante un estudio analítico descriptivo, se identificó previa selección los buscadores que operan en la red el grado de efectividad con la cual las herramientas de búsqueda recuperan la información y la relación entre la recuperación de la información y el uso de metadatos.

Palabras clave: recuperación de información; herramientas de búsqueda; motores de búsqueda web; información electrónica.

Recibido: 11/1/17 Devuelto para revisión: 23/1/17 Aceptado: 20/2/17

¹ Profesor de tiempo completo en el Colegio de Bibliotecología de la UNAM. Estudiante del Doctorado en Bibliotecología y Estudios de la Información de la Universidad Nacional Autónoma de México. Maestría en Bibliotecología, FFyL-UNAM. Licenciatura en Bibliotecología, FFyL-UNAM. Correo electrónico: amay@unam.mx

Abstract

The exponential growth in the volumes of information available through the internet, leads to the users of the network of networks, use the search tools, a situation that leads to face several problems at the time of recovering the required contents. Based on this, it is necessary to evaluate the performance of different tools of search and retrieval of the information found on the web. Using a descriptive analytical study was identified after selecting the search engines that operate in the the degree of effectiveness with which the search tools retrieve information and the relationship between the recovery of the information and the use of metadata.

Keywords: information retrieval; search tools; web search engines; electronic information.

Introducción

En la denominada sociedad de información, se concibe este recurso como estratégico y logra consolidarse en la actualidad con alta importancia, quizás superior a la acreditada a otros bienes en otras épocas. Parte de lo planteado, se acredita al avance de la tecnología, la cual se ve reflejado en muchos ámbitos de la cotidianidad hasta las formas de enseñanza. De tal manera, que la información no ha sido ajena a este avance tecnológico, ni puede quedarse al margen de ella, esto se ve reflejado en la explosión de información en la red conocida como Internet, parte integrante de la concepción de las tecnologías de la información.

Las tecnologías de información, han producido cambios insospechados respecto a cómo se produce y se utiliza la información en la actualidad. La información, como uno de los activos principales de las organizaciones y labor profesional, ha sufrido cambios tanto en la forma de generarse y distribuirse como en la que se obtiene.

El avance tecnológico al que se hace referencia, se generó a finales del siglo XX y se proyecta su crecimiento y desarrollo en el presente, mediante nuevas formas de crear información, principalmente apoyadas en medios electrónicos, haciendo posible que casi cualquier persona pueda crear y publicar en Internet. Tal situación, ha ocasionado que se incremente de manera incontrolable la información disponible en la red, lo cual trae como consecuencia que los usuarios se enfrenten a una red con diversidad de problemas entre los cuales se mencionan como poder acceder de forma directa y organizada, así como también poder recuperar la información que sea relevante y de interés ante las demandas y necesidades de los usuarios.

Entre las causas más comunes a las que se enfrenta el usuario, al tratar de recuperar documentos electrónicos en las herramientas de búsqueda que se encuentran en la red, están las siguientes: la falta de datos suficientes para su indización, inadecuado uso de estructuras de metadatos y el uso de protocolos inadecuados. Razones por la que han surgido diferentes propuestas para mejorar las formas de organización de la información, con el fin de recuperar eficientemente la información de la red.

En función de lo expuesto, se han considerado los metadatos, como un mecanismo mediante la cual se organiza la forma en que un sitio es indexado por los motores de búsqueda. Es importante mencionar, que en la actualidad la mayoría de los principales buscadores utilizan metadatos para construir los índices de sus bases de datos. De este modo, el objetivo de esta investigación es el de evaluar el funcionamiento de diferentes herramientas de búsqueda y recuperación de la información disponible en la Web, para identificar si existe relación entre la recuperación de la información y el uso de metadatos.

Para su desarrollo, se parte de los siguientes supuestos: si se utiliza una estructura de metadatos adecuada en la organización de la información, se obtiene un mayor índice de recuperación de la información. Sin embargo, muchas de las herramientas de búsqueda de información de la web no cuentan con estructuras de metadatos adecuadas, lo que dificulta la recuperación de la información, en particular la recuperación temática.

Explosión de la información en la Web

La red de redes Internet, en particular la web está viviendo su propio Big Bang. Año con año, la numerosa información que cerca de mil 100 millones de usuarios suben a las redes sociales, sitios web y otros se multiplica considerablemente, marcando un crecimiento nunca antes visto. Es decir, que la información de la web crece el 60% al año, del cual los usuarios aportan el 75% de lo publicado. (Gantz, John; Reinsel David, 2010) El estudio “Extracting Value from Chaos” se menciona que es tanta la información que existe en la web, que ya no es posible medirla en *Gigabytes*: ahora la medida va en *Zettabytes* (es decir, un millón de millones de Gigabytes). (Gantz, John; Reinsel David 2011)

Durante el 2010, la cantidad de datos superó holgadamente el zettabyte, y se espera que año con año la información alcance los 1,8 ZB. Para el 2020, esta información debiera alcanzar 35,2 ZB, producto de múltiples factores que van aportando a este crecimiento: las producciones de los usuarios, tales como fotos, correo y variedad de escritos, que han ido migrando hacia sus versiones digitales, destronando en ciertos casos a los formatos en papel. Si a eso sumamos la expansión de las nuevas formas de comunicación –como mensajería instantánea, redes sociales y blogs- y las herramientas para compartir, copiar y replicar contenidos, se puede explicar buena parte de dicho crecimiento. (Gantz, John; Reinsel David 2011)

Según estudio realizado por Nielsen Online* en 2010 se presenta como aumentó el número de sitios web con 47 millones en el año 2009, asimismo (blog pulse, 2010) menciona que en 2009 había aproximadamente 126 millones de blogs en Internet. Con respecto a fotos y videos, hasta octubre del 2009 Flickr tenía alojadas 4,000,000,000 fotos y en el caso de Facebook se suben 2,500,000,000 de fotos al mes y 30 mil millones de fotos al año. En el caso de los videos pasa algo completamente diferente a las fotos, esto debido a que es muy difícil calcular el número de videos que existen en la red, como es el caso de YouTube que en un día se consultan aproximadamente mil millones de videos y al mes se ven 12,200 millones de videos.

Lo antes expuesto, permite conocer desde una aproximación en la actualidad la cantidad de información que existe en Internet. Sin embargo, aún más impresionante es la presentación de la siguiente información tomada de una infografía llamada Data Never Sleeps How much Data is Generated Every Minute, creada por el personal de DOMO junto a Column Five Media que se enfoca principalmente en la gigantesca cantidad de información que se mueve cada minuto en Internet. (Domo.com 2012)



Fuente:Data Never Sleeps Infographic, DOMO (2012).

La infografía presentada, permite evidenciar los datos del crecimiento de Internet minuto a minuto. Según se especifica:

- La Web Móvil recibe 217 nuevos usuarios.
- Los usuarios de YouTube suben 48 horas de nuevos videos.
- Se envían 204.166.667 de correos electrónicos.
- Google recibe 2.000.000 de búsquedas.
- Los usuarios de Facebook comparten 684.478 piezas de contenido.
- Los usuarios de WordPress publican 347 nuevos posts.
- Son creados 571 nuevos sitios.
- Los usuarios de Foursquare realizan 2.083 check-ins.
- Los usuarios de Instagram comparten 3.600 nuevas fotos.
- Apple recibe 47.000 descargas de aplicaciones.
- Los usuarios de Tumblr publican 27.778 nuevos posts.
- Los usuarios de Flickr agregan 3.125 nuevas imágenes.
- Las organizaciones en Facebook reciben 64.722 nuevos “Me Gusta”.
- Se envían 100.000 tweets en Twitter.
- Se gastan U\$S272.070 dólares en compras por Internet.

Lo descrito ha ocasionado, el incremento alarmante de la información disponible en la red. Situación que trae como consecuencia, que los usuarios se enfrenten a una red con muchos problemas y carente de organización para poder recuperar la información que sea relevante para sus necesidades. Asimismo, la transformación de lo impreso a medios electrónicos es, sin duda alguna, muestra de los cambios que se evidencian en nuestra época. El uso de tecnologías de información, como el caso de la Internet en la generación, la búsqueda, distribución y recuperación de información abre una nueva posibilidad dentro del ámbito bibliotecario.

Metadatos y recuperación de información

En la actualidad uno de los medios de alta preferencia es la web, requerida para buscar e investigar sobre un tema en particular, de allí la importancia de conocer cómo funciona y las alternativas para logara el efectiva recuperación de información en la Red. Si bien es cierto, que la web de altos volúmenes de volúmenes de información, también dispone de diferentes herramientas que nos permiten recuperar la información

pertinente para satisfacer las necesidades de información de los usuarios. El proceso de búsquedas de información en la Red, resulta para los usuarios en frecuentes oportunidades dos problemas, el primero es el exceso de información y en ocasiones es muy difícil separar las páginas que disponen de soportes académicos con respecto a los que no lo tienen.

El segundo problema, son las debilidades para recuperar documentos con las características requeridas, esto ocasiona la pérdida de información en la red, y se considera como una de las causas que los recursos de información son almacenados en el web, por sus propios creadores, sin ningún tipo de descripción que permita su organización y posterior recuperación, y esto posiblemente sea uno de los puntos débiles de la red. En consecuencia, son incalculables los documentos que no pueden ser recuperados por no contar con una adecuada descripción de sus características que le permita a las herramientas de búsqueda su localización y recuperación.

Por esta razón los metadatos, surgen como un punto de apoyo para ayudar a organizar y posteriormente recuperar la información. Caplan (1995) menciona que en el ámbito bibliotecario, el término metadatos fue acuñado por Jack Myers en la década de los 60 para describir conjuntos de datos. La primera acepción que se le dio (y actualmente la más extendida) fue la de dato sobre el dato, ya que proporcionaban la información mínima necesaria para identificar un recurso.

En el sentido actual, los metadatos son datos sobre los datos, esto es metaelementos sobre el contenido y el contexto de los documentos digitales que permiten marcar los campos utilizados por los robots de búsqueda para clasificar y encontrar la información. Descriptor básico de recursos que facilita su identificación y recuperación. (Pinto y Lancaster 2001)

Por lo tanto, los metadatos deben contener información estructurada para que las máquinas las entiendan, en este sentido los metadatos son datos altamente estructurados que describen información, contenido, la calidad, proceso de producción de datos, detalles de formatos y otras características de los datos. Las estructuras de metadatos se encuentran conformadas por campos, lo que permite su fácil utilización en la organización y recuperación de la información.

Dentro del ámbito de las herramientas de búsqueda en la Web, además de crearlas y utilizarlas, es necesario conocer cuál es la forma en la que están trabajando, para saber si cumplen con su objetivo principal, el de permitir que los usuarios recuperen la información que necesitan en pro de la satisfacción de las demandas presentadas. Para conocer si estas herramientas recuperan adecuadamente la información, es necesario realizar un estudio, consistente en una búsqueda de información utilizando una estrategia, que nos permita conocer el grado de efectividad con que las herramientas de búsqueda recuperan la información y que tipo de metadatos utilizan.

Herramientas de búsqueda de información

Para organizar la información y recuperarla se han desarrollado herramientas de búsqueda, orientadas hacia la posibilidad de garantizar la recuperación rápida, efectiva y eficiente de la información. Como es

sabido, prevalecen un sin número de documentos que pueden ser accedidos mediante las herramientas de recuperación de información que existen, entre los cuales se mencionan los siguientes tipos de buscadores, entre los que se encuentran: *directorios, motores de búsqueda, metabuscadores, multibuscadores, agentes inteligentes de búsqueda de información y buscadores semánticos*

Las opciones mencionadas, presentan diferentes características, lo cual ha ocasionado que los usuarios no conozcan qué tipo de herramienta es la que están utilizando y por lo tanto usan el término de buscador o motor de búsqueda de manera indistinta. Por esta razón, de forma breve se presenta la conceptualización de cada una de las herramientas mencionadas

Buscadores: en la década de los 90s, los buscadores tuvieron un desarrollo impresionante. Desde los primeros días de la web, estos se han convertido en una herramienta indispensable para los usuarios de Internet. Por lo tanto, son los que contienen los índices de la información que se encuentra en la Web, además de ser necesarios para los usuarios que buscan y crean recursos de información, así como también puedan utilizar estos índices que se encuentran almacenados en cada buscador.

El buscador ha sido definido como un programa informático diseñado para la búsqueda de diferentes formatos de archivos digitales como páginas web, documentos de texto, archivos de video y música, imágenes. Los cuales se encuentran alojados en un disco duro o servidor local, así como en la red de Internet, presentando los resultados de la búsqueda en una serie de enlaces que conectan con los archivos o la información solicitada. (Qué es un buscador.2016)

En términos generales, un buscador es un programa informático, que permite consultar una base de datos en la cual se encuentran los índices que relacionan direcciones de páginas web con su contenido. Su uso facilita enormemente la obtención de un listado de páginas web que contienen información sobre algún tema en particular www.quees.info, (2016).

Asimismo, Levene 2000 mencionan que existe una gran cantidad de buscadores, clasificados en tres grandes rubros, como ya se mencionó en párrafos anteriores: directorios, motores de búsqueda, metabuscadores. Estos tres tipos de buscadores, son los más utilizados por los usuarios que requiere hacer búsquedas de información en la WWW, asimismo cada uno de ellos cuenta con sus características propias, ventajas y desventajas en la recuperación de información.

Directorios

Los directorios fueron las primeras herramientas de recuperación de información utilizadas en Internet, su funcionamiento tiene que ver con el trabajo humano, principalmente bibliotecarios, que indexan en una base de datos, una serie de datos relevantes sobre un recurso de información, como son: el título, resumen del contenido y la dirección electrónica o URL. Según, el sitio www.quees.info, (2016) todo esto, dividido en categorías, con el objetivo de garantizar al usuario la efectividad de las búsquedas de información, la base de datos le devolverá una lista de enlaces hacia las páginas webs que tienen relación con las palabras o frases buscadas.

Los directorios, son índices temáticos en los que se pretende clasificar los sitios web mediante categorías temáticas jerarquizadas. Generalmente son indexados por personal bibliotecario, aunque en la actualidad, existen algunos directorios como Yahoo, que iniciaron su indización de manera manual, pero con el incremento de información en la red, migraron su forma de indizar la información y actualmente conservan su estructura de directorio pero ahora utilizan los robots de Google y Bing para indizar automáticamente los recursos de información en su base de datos, www.quees.info, (2016).

De igual manera, son consideradas "guías estructuradas y sistematizada por clases y subclases reflejan diferentes direcciones y descripciones por diversas áreas del conocimiento de acuerdo a la fuente", Escobar, Paredes y Barroso, (2015). Los directorios, son útiles para los usuarios que acceden a la red sin una clara orientación de cómo buscar información. Dentro de las características que podemos mencionar de estos directorios se encuentran las siguientes: contruidos por selección humana. (en algunos casos), organizados por categorías temáticas, alta calidad en contenido de los sitios web indizados, los hay de diferentes tamaños desde pequeños hasta más amplios y especializados, pero siempre más pequeños que los motores de búsqueda.

Motores de búsqueda

Estas herramientas de búsqueda, son las más comunes en la Red. Los motores de búsqueda son un tipo de buscador basado en programas informáticos que automatizan el rastreo y la indexación de las páginas web disponibles, estos surgen de la necesidad de organizar la información anárquica contenida en Internet, a diferencia de los directorios los cuales necesitan de una parte humana para la verificación e indexación de las páginas web. El funcionamiento de los motores de búsqueda se basa en 3 etapas o fases: rastreo, indexación y ranqueo y consulta, www.quees.info, (2016).

En la primera fase un programa informático denominado robot o araña rastrea y recoge los datos de todos los recursos de información en el Web, a través de los enlaces compartidos, permitiendo ampliar enormemente el campo de rastreo a un gran número de información. En el segundo, paso un algoritmo toma la información captada por el robot o araña y la valora rankeandola según criterios establecidos por el algoritmo, posteriormente se indexa la información en gigantescas bases de datos. Por último, www.quees.info, (2016) cuando se realiza una búsqueda el programa consulta a la base de datos mostrando una serie de enlaces o datos numerados y jerarquizados según el ranqueo o relevancia calculada por el algoritmo.

Cuando un motor efectúa una búsqueda, la información es rastreada en el contenido de una base de datos (no se busca directamente en el World Wide Web), por lo que ninguno de estos buscadores incluye en su base de datos todos los índices de la información existente en el WWW, por lo tanto se obtienen resultados diferentes.

Dentro de las principales características que podemos mencionar de estos buscadores se encuentran las siguientes: construyen sus bases de datos con la ayuda de programas llamados spiders, robots, no están organizados por temas, no son evaluados, contienen páginas de todo tipo y sus bases de datos son enormes.

Es importante mencionar, que las diferencias en los resultados obtenidos en las búsquedas son debido a las características particulares de cada uno de los buscadores.

Metabuscadore

Con la continua aparición de diferentes herramientas y para agilizar la búsqueda y recuperación de información en internet, han aparecido otras herramientas de búsqueda de información en la red, como es el caso de los metabuscadores. Estos son programas informáticos que permiten la realización de una misma búsqueda en diversos buscadores de forma simultánea, su funcionamiento se basa en lanzar las consultas requeridas por los usuarios sobre un conjunto de motores de búsqueda y directorios, aprovechando los algoritmos de búsqueda desarrollados por cada uno de ellos, posteriormente el metabuscador rankea todos los resultados mostrándolos según criterios de relevancia establecidos, su interés radica en la reducción del tiempo de búsqueda por la simultaneidad de herramientas que revisa.

Dentro de las principales características de los metabuscadores encontramos las siguientes: eliminan los resultados duplicados, eliminan los links rotos, o que no están disponibles, utilizan la lógica de recuperación de otros buscadores y no cuentan con base de datos propia. Estos metabuscadores, cada vez se han ido mejorando, en la actualidad algunos trabajan por medio de clúster (clustering), que permite reducir el volumen de información mediante la categorización o agrupamiento de los datos con características similares, la cual es una tarea natural del intelecto humano, www.quees.info, (2016).

El clustering, aprovecha esta característica aportando herramientas que faciliten la construcción automática de taxonomías y minimizando la intervención humana en este proceso. Este tipo de clasificación automática es conocida también como taxonomía numérica, debido a que fue desarrollada originalmente para ser aplicada en clasificaciones biológicas. Existen alrededor de 150 técnicas diferentes de clustering. En documentación se han utilizado sobre todo técnicas jerárquicas. (Pinto 2005)

Multibuscadores

Al existir una gran cantidad de diferentes buscadores en el web, y para ayudar a los usuarios a consultar en una sola pantalla, sin salir de ella, diferentes directorios, motores de búsqueda y metabuscadores, aparecieron los llamados multibuscadores, ellos se limitan a colocar en una página *web* una lista con las herramientas más populares y con una ventana de texto para poder enviar la cadena de búsqueda introducida por el usuario a cada uno de ellas. Al usar un multibuscador se envía la cadena de búsqueda a una serie de motores de búsqueda y directorios previamente definidos.

Una vez realizada la consulta a los buscadores prefijados, el multibuscador devolverá la información de todos los enlaces encontrados. Con la utilización de multibuscadores, el usuario se evita tener que ir de buscador en buscador hasta encontrar la información deseada. El usuario elige los buscadores que quiere utilizar y desde

allí realiza su consulta que ofrecerá las respuestas buscador por buscador. La única ventaja, es la posibilidad de consultar un gran número de buscadores partiendo de una única página. La exhaustividad privará sobre la precisión, ya que el usuario encontrará un gran número de enlaces y muchas páginas repetidas [www.quees.info](#), (2016).

Agentes inteligentes de búsqueda de información

En párrafos anteriores se ha mencionado del gran crecimiento de internet y de la proliferación de buscadores y metabuscadores, como una ayuda para la recuperación de la información, a estos se han sumado otras herramientas de búsqueda denominadas agentes inteligentes de búsqueda de información. Estas herramientas constituyen la evolución lógica de los metabuscadores, debido a que buscan también en varios motores de búsqueda y directorios, solo que para su utilización es necesario la instalación de softwares clientes.

Dentro de las principales características de los agentes inteligentes encontramos las siguientes: eliminan los documentos duplicados, permiten el trabajo offline, esto es, sin estar conectados a internet, permiten editar y guardar las estrategias de búsqueda

Buscadores semánticos

Este tipo de buscadores, se pueden considerar como la evolución en la recuperación de información, es decir, son las herramientas del futuro, para llegar no propiamente a la Web semántica, sino a una técnica semántica de recuperar la información en cualquier sistema de información.

Un buscador semántico deberá efectuar la búsqueda atendiendo al significado del grupo de palabras que se solicita en una búsqueda. Los buscadores semánticos, interpretaran las búsquedas de los usuarios y los textos de los documentos en la web a través del uso de algoritmos que representan comprensión o entendimiento, ofreciendo resultados correctos de una manera rápida, sin que el usuario tenga que abrir los documentos e inspeccionarlos. De esta manera, los buscadores semánticos reconocen el contexto de las palabras o sentencias de búsqueda. Con los buscadores semánticos, se persigue que la búsqueda que realice el usuario sea más precisa y origine resultados relevantes con la mínima intervención de este.

En función de lo expuesto, se mencionan los elementos necesarios para que este este tipo de recursos pueda realizar búsquedas semánticas: en primer lugar necesitan utilizar metadatos codificados, es decir, que deben utilizar un esquema de codificación que podría ser el formato MARC o bien un lenguaje de marcado como el XML, los cuales permitirían codificar los metadatos. En segunda instancia, es necesario que cuente con un esquema que le permita dotar a los documentos de un significado de relaciones entre diferentes aspectos de un documento, como podría ser qué relación tiene el tipo de recurso, con el título y el autor, en este sentido el esquema que podría solucionar los aspectos antes mencionados, es el RDF (Marco de Descripción de Recursos) Resource Description Framework (RDF), (2014).

Y finalmente debe contener también ontologías, las que introducirán un mayor grado de profundización semántica y proporcionaran una descripción lógica y formal que pueda ser interpretada, tanto por las personas, como por las máquinas, asimismo las ontologías permiten la interoperabilidad entre distintos sistemas.

En la actualidad existe una serie de buscadores llamados semánticos, sin embargo es necesario evaluarlos para saber si reúnen las características que debe de tener un buscador, para realizar búsquedas con una técnica semántica apropiada para recuperar eficientemente la información. La creación y uso de estos buscadores en la Web todavía es incipiente, pero se siguen haciendo esfuerzos para llegar a tener herramientas que ayuden de manera importante a los usuarios de la red a recuperar la información de manera más eficiente.

Metodología de la Evaluación

Para conocer si las herramientas previamente desarrolladas, recuperan adecuadamente la información, se llevó a cabo un estudio, que permitiese conocer el grado de efectividad con que las herramientas de búsqueda recuperan la información. Para tales efectos, se realizó una selección de los buscadores que se encuentran operando en la red. Lo cual resultó difícil, debido a que existe una gran cantidad de diferentes tipos y tamaños. Por lo cual, se utilizaron solo los motores de búsqueda, de mayor frecuencia de uso en la Web, previa identificación del ranqueo correspondiente, de los cuales se mencionan los tres principales: eBizMBA(2016), ListofSearchEngines (2016) y Reliablesoft (2016).

Con respecto a los metabuscadores, se tomó como referencia para su selección a los que aparecen en la página de ListofSearchEngines, en el rubro de metabuscadores y por ser los más utilizados. Para la selección de los buscadores semánticos, se utilizaron dos sitios Web: Search engine journal (2016) y Makeuseof(2016).

Es importante aclarar que solo se utilizaron 4 buscadores semánticos por ser los únicos disponibles. Como resultado de este proceso de selección de herramientas de búsqueda de información, se consideraron cinco motores de búsqueda, cinco metabuscadores y cuatro buscadores semánticos ver tabla 1, estos fueron:

Tabla 1. Motores de Búsquedas, metabuscadores, buscadores semánticos

Motores de búsqueda	Metabuscadores	Buscadores semánticos
Google	WebCrawler	Kngrine
Yahoo	Dogpile	Exalead
Bing	Info.com	DuckDuckgo
Ask	Zapmeta	Swoogle
AOL	Exite	

Fuente: eBizMBA, Listofsearchengnes, Reliablesof.

Finalmente fue necesario estructurar una estrategia de búsqueda, que nos permitiera conocer si las herramientas de búsqueda recuperaban la información eficientemente. Como otro aspecto importante, en la investigación fue la selección del tema para efectuar las búsquedas en las diferentes herramientas, el tema seleccionado, fue:

En español: Las relaciones diplomáticas entre Estados Unidos y Cuba

En inglés: Diplomatic relations between the United States and Cuba

La decisión de incluir este tema obedeció a que es una temática muy relevante en estos momentos, tanto para los dos países involucrados, Estados Unidos y Cuba, así como para el ámbito internacional. Por otro lado, se decidió que la búsqueda debería realizarse en idioma Inglés y español, debido a que son los idiomas oficiales de los países protagonistas. Por lo que, se procedió a realizar tres tipos de búsqueda:

Búsqueda libre, donde se incluían todos los términos:

Ejem. español: **relaciones diplomáticas entre estados unidos cuba**

Ejem. Inglés: **diplomatic relations between united states cuba**

Búsqueda utilizando operadores lógicos boléanos, solo operador AND y comillas.

Ejem. español: **relaciones AND diplomáticas AND entre AND “estados**

unidos” AND cuba
relations AND between AND “united states”

Ejem. Inglés: **diplomatic AND**

AND cuba

Búsqueda utilizando operadores lógicos en

el campo de título.

Ejem. español: **intitle:relaciones AND intitle:diplomáticas AND intitle:entre**

AND intitle:“estados unidos” AND intitle:cuba
AND intitle:relations AND intitle:between AND

Ejem. Inglés: **intitle:diplomatic**

intitle: “united states” AND intitle:cuba

Resultados

La búsqueda de la información antes mencionada, en los diferentes motores de búsqueda, metabuscadores y buscadores semánticos, seleccionados para este estudio, permiten observar que en la búsqueda libre en español, los motores de búsqueda recuperaron enormes cantidades de información.

Entre los motores que más resultados ofrecieron se encuentran, en primer lugar el motor Bing con 2,660,000, en segundo lugar AOL con 608,000, en tercero Google con 520,000, en cuarto lugar Yahoo con 145,000 y en quinto lugar ASK con solo 100 resultados.(véase tabla 2)

En la búsqueda utilizando operadores lógicos los resultados fueron similares, en primer lugar fue el motor Bing con 2.290,000 resultados, en segundo lugar AOL con 462,000, en tercer lugar Google con 457,000, en cuarto Yahoo con 129,000 y en quinto lugar ASK con 100 resultados. Los resultados en la búsqueda utilizando el campo de título quedaron de la siguiente manera, en primer lugar el motor Yahoo con 5,720 resultados, en segundo Bing con 1,860, en tercero Google con 14, los motores ASK y AOL no obtuvieron resultados.(véase tabla 2)

Tabla 2 Resultados de los motores de búsqueda en español

Motor	Libre	Operadores lógicos	Campo de titulo
Google	520, 000	457, 000	14
Yahoo	145, 000	129, 000	5, 720
Bing	2, 660,000	2, 290,000	1,860
Ask	100	100	0
AOL	608,000	462,000	0

Fuente: elaboración propia, (2016).

Con respecto a los resultados obtenidos a través de los cinco motores, en las búsquedas en español de forma libre, utilizando los operadores lógicos boléanos y utilizando el campo de título, fueron los siguientes ver tabla 3:

Tabla 3. Resultados de los motores de búsqueda

Google	Yahoo	Bing	Ask	AOL
L= 331	L= 183	L= 380	L= 100	L= 203
O= 331	O= 189	O= 423	O= 100	O= 186
T= 13	T= 121	T= 389	T= 0	T= 0

Fuente: elaboración propia , (2016).

Con respecto a las búsquedas en español en los metabuscadores, los resultados fueron los siguientes permitieron determinar que en la búsqueda libre en primer lugar, el metabuscador que más resultados obtuvo fue Zapmeta con 1,650,000, en segundo lugar Dogpile con 940, en tercero WebCrawler con 730, en cuarto Exite con 400 y en quinto Info.com con 300 resultados. La búsqueda con operadores lógicos, en primer lugar con más resultados apareció Zapmeta con 33,800 resultados, en segundo WebCrawler con 670, en tercero Exite con 400, en cuarto Dogpile con 370 y en quinto con 320 resultados. La Búsqueda en el campo de título, no obtuvo resultados en ningún metabuscador.(véase tabla 4)

Tabla 4 Resultados de los metabuscadores en español

Metabuscadores	Libre	Operadores lógicos	Campo de título
WebCrawler	730	670	0
Dogpile	940	370	0
Zapmeta	1, 650,000	33,800	0
Info.com	300	320	0
Exite	400	400	0

Fuente: elaboración propia, (2016).

Los resultados que se pudieron visualizar en los cinco metabuscadores en las búsquedas en español, de forma libre, con operadores lógicos booleanos y utilizando el campo de título, fueron los siguientes ver tabla 5:

Tabla 5. Resultados visibles de los metabuscadores en español

WebCrawler	Dogpile	Zapmeta	Info.com	Exite
L= 730	L= 940	L= 50	L= 300	L= 400
O= 670	O= 370	O= 50	O= 320	O= 400
T= 0	T= 0	T= 0	T= 0	T= 0

Fuente: elaboración propia, (2016).

Los resultados de las búsquedas en español, en los buscadores semánticos fueron los siguientes, con más resultados apareció en primer lugar Exalead con 40,790, en segundo DuckDuckgo con 100, en tercero Kngine con 15 resultados y el buscador Swoogle no recuperó ningún resultado. En la búsqueda con operadores lógicos, con más resultados en primer lugar apareció Exalead con 44,297 resultados, en segundo DuckDuckgo con 100, en tercero Kngine con 15 resultados y Swoogle no obtuvo resultados. En las búsquedas en el campo de

título, solamente el buscador semántico Exalead recuperó 41 resultados, los otros buscadores no obtuvieron ningún resultado.(véase tabla 6)

Tabla 6 Resultados de buscadores semánticos en español

Buscadores Semánticos	Libre	Operadores lógicos	Campo de titulo
Kngine	15	15	0
Exalead	40,790	44,297	41
DuckDuckgo	100	100	0
Swoogle	0	0	0

Fuente: elaboración propia, (2016).

Los resultados que se pudieron visualizar en los cuatro buscadores semánticos, utilizando los tres tipos de búsquedas en español, fueron los siguientes ver tabla 7:

Tabla 7. Resultados visibles de buscadores semánticos en español

Kngine	Exaleade	DuckDuckgo	Swoogle
L= 15	L= 40,790	L= 100	L= 0
O= 15	O= 44,297	O= 100	O=0
T= 0	T= 41	T= 0	T= 0

Fuente: elaboración propia, (2016).

Los resultados relacionados con la búsqueda libre en inglés, permitieron la recuperación de enormes cantidades de información, entre los motores que más resultados ofrecieron se encuentran, primer lugar el motor AOL con 14, 300,000, en segundo lugar apareció Google con 10.900,000, en tercero Bing con 9, 960,000, en cuarto lugar Yahoo con 7, 270, 000 y en quinto lugar ASK con solo 100 resultados. En la búsqueda utilizando operadores lógicos, los resultados fueron similares, en primer lugar fue el motor AOL con 12, 400,000 resultados, en segundo lugar Google con 10,100,000, en tercer lugar Bing con 4, 140,000, en cuarto Yahoo con 606, 000 y en quinto lugar ASK con 100 resultados.(véase tabla 7)

Los resultados en la búsqueda utilizando el campo de título quedaron de la siguiente manera, en primer lugar el motor Yahoo con 46, 800 resultados, en segundo Bing con 128, en tercero Google con 63, los motores ASK y AOL no obtuvieron resultados. (véase tabla 8)

Tabla 8 Resultados de los motores de búsqueda en inglés

Motor	Libre	Operadores lógicos	Campo de titulo
Google	10,900,000	10,100,000	63
Yahoo	7, 270, 000	606, 000	46, 800
Bing	9, 960,000	4, 140,000	128
Ask	100	100	0
AOL	14, 300,000	12, 400,000	0

Fuente: elaboración propia, (2016).

Con respecto a los resultados que visualizados en los cinco motores, utilizando los tres tipos de búsquedas en inglés, fueron los siguientes ver tabla 9:

Tabla 9. Resultados visibles de los motores de búsqueda en inglés

Google	Yahoo	Bing	Ask	AOL
L= 352	L= 190	L= 445	L= 100	L= 192
O= 352	O= 200	O= 441	O= 100	O= 194
T= 27	T= 222	T= 128	T= 0	T= 0

Fuente: elaboración propia, (2016).

Con respecto a las búsquedas en inglés en los metabuscadores, los resultados fueron los siguientes: en la búsqueda libre en primer lugar, el metabuscador que más resultados obtuvo fue Zapmeta con 39.600,000, en segundo lugar WebCrawler con 430, en tercero Dogpile con 410, en cuarto Info.com con 400 y en quinto Exite sin resultados. La búsqueda con operadores lógicos, en primer lugar con más resultados apareció Zapmeta con 1, 230,000 resultados, en segundo Exite 460, en tercero Info.com con 380, en cuarto WebCrawler con 300 y

en quinto Dogpile con 10 resultados. La Búsqueda en el campo de título, Dogpile con 130 resultados, Exite con 100, Info.com con 10 y Zapmeta y WebCrawler no se obtuvieron resultados. (véase tabla 10)

Tabla 10 Resultados de los metabuscadores en inglés

Metabuscadores	Libre	Operadores lógicos	Campo de titulo
WebCrawler	430	300	0
Dogpile	410	10	130
Zapmeta	39,600,000	1,230,000	0
Info.com	400	380	10
Exite	0	460	100

Fuente: elaboración propia, (2016).

En el caso de los resultados que se pudieron visualizar en los cinco metabuscadores, utilizando los tres tipos de búsqueda en inglés, fueron ver tabla 11:

Tabla 11. Resultados visibles de los metabuscadores en inglés

WebCrawler	Dogpile	Zapmeta	Info.com	Exite
L= 430	L= 410	L= 50	L= 400	L= 0
O= 300	O= 10	O= 50	O= 380	O= 460
T= 0	T= 130	T= 0	T= 10	T= 100

Fuente: elaboración propia, (2016).

Los resultados de las búsquedas en inglés, en los buscadores semánticos fueron los siguientes, con mayor resultados apareció en primer lugar Exalead con 153,277, en segundo DuckDuckgo con 100, en tercero Kngine con 15 resultados y el buscador Swoogle no recuperó ningún resultado. En la búsqueda con operadores lógicos, con más resultados en primer lugar apareció Exalead con 94,753 resultados, en segundo DuckDuckgo con 100, en tercero Kngine con 15 resultados y Swoogle no obtuvo resultados. En las búsquedas en el campo de título, solamente el buscador semántico Exalead recuperó 10 resultados, los otros buscadores no obtuvieron ningún resultado. (véase tabla 12).

Tabla 12. Resultados de las búsquedas en inglés, en los buscadores semánticos.

Buscadores Semánticos	Libre	Operadores lógicos	Campo de titulo
Kngine	15	15	0
Exalead	153,277	94,753	10
DuckDuckgo	100	100	0
Swoogle	0	0	0

Fuente: elaboración propia, (2016).

Los resultados visualizados en los cuatro buscadores semánticos, utilizando los tres tipos de búsqueda, quedando de la siguiente manera ver tabla 13:

Tabla 13. Resultados visibles de las búsquedas en inglés, en los buscadores semánticos.

Kngine	Exaleade	DuckDuckgo	Swoogle
L= 15	L= 153,277	L= 100	L= 0
O= 15	O= 94,753	O= 100	O=0
T= 0	T= 10	T= 0	T= 0

Fuente: elaboración propia, (2016).

Análisis de los resultados

A través de los resultados obtenidos, se observa que los motores de búsqueda, en los diferentes ejemplos presentados tanto en español como en inglés, presentaron diferentes resultados, en la búsqueda libre cuatro de ellos , recuperaron enormes cantidades de información y solamente uno recupero 100 resultados, como parte de la interpretación se considera que entre más términos se incluyan de forma libre en una búsqueda más resultados de obtienen, debido a que las bases de datos de los motores de búsqueda tienen asignado por default el operador lógico OR, de tal manera que cada termino que se anexa a la búsqueda recupera más

información que no es tan precisa, por lo tanto las búsquedas de este tipo recuperan más información que no tiene nada que ver con la verdadera búsqueda.

En el caso de la búsqueda utilizando las comillas y el operador lógico AND presentó en cuatro motores una reducción importante con respecto a los resultados obtenidos en la búsqueda libre y solamente un motor recuperó 100 resultados, lo cual indica que este último motor no trabaja de la misma manera que los otros cuatro, debido a que al utilizar los operadores lógicos booleanos, la mayoría de los motores garantizan que todos los términos de la búsqueda aparecen en el mismo documento y por lo tanto el número de resultados se reduce de manera importante.

Con respecto a la búsqueda en el campo de título, solo tres motores recuperaron información y los otros dos no recuperaron ningún documento, tal situación permite pensar que estos motores al recuperar información en el campo de título, sus bases de datos si utilizan metadatos para la organización de los índices de información que están almacenados en sus bases de datos, pero definitivamente los metadatos que utilizan no son los adecuados para organizar la información y por lo tanto la recuperación de información no es tan precisa.

Otro aspecto importante en estas búsquedas es que al tratar de visualizar los documentos que recuperaron los diferentes motores, solamente se podía tener accesibilidad a una cantidad muy pequeña de los que supuestamente recuperó el motor, por lo tanto esto reafirma que los metadatos que utilizan estos motores de búsqueda no son los adecuados y por lo tanto no se puede revisar todos los documentos obtenidos, o también se puede pensar que los motores presentan el número de veces que un término aparece en sus índices y no el número de documentos que existen indizados en sus bases de datos.

En el caso de los resultados que se obtuvieron en los metabuscadores, tanto en español como en inglés, en la búsqueda libre, se puede mencionar que existe una reducción en la recuperación de documentos en la mayoría de los metabuscadores, aunque Zapmeta recuperó una gran cantidad de documentos, y los otros cuatro recuperaron menos de 1000 cada uno, por lo tanto se observa que los metabuscadores al recuperar la información en diferentes bases de datos de los motores de los cuales tienen autorización para buscar en sus índices, hacen una selección de los mejores resultados, es decir buscan y recuperan los documentos más relevantes en cada búsqueda y los presentan al usuario.

Al realizar la búsqueda utilizando operadores lógicos booleanos en estos metabuscadores, se observa que en realidad si hacen una reducción de resultados, pero aun así, no existe una precisión en la recuperación de información.

Por otro lado al realizar la búsqueda en el campo de título no se obtuvieron resultados en ningún metabuscador, por lo tanto esto indica que no utilizan metadatos, por obvia razón, al utilizar la estructura de búsqueda de los motores y no contar con bases de datos propias, solamente presentan los resultados que recuperan en éstos motores de búsqueda.

En el caso de los buscadores semánticos en la búsqueda libre, dos buscadores recuperaron la misma cantidad de resultados, tanto en la búsqueda libre como en la búsqueda con operadores lógicos booleanos, en la búsqueda

en el campo de título no obtuvieron resultados, el buscador Swoogle, no recuperó ningún resultado en las tres búsquedas y solamente un buscador recuperó una gran cantidad de resultados en la búsqueda libre y con la utilización de operadores también se vio una reducción de documentos, finalmente en la búsqueda por título también recupero información.

Situación que conlleva a pensar que si de cuatro buscadores semánticos solamente uno trabaja de forma adecuada los otros tres no utilizan alguna estructura de metadatos en sus bases de datos para organizar y recuperar la información, el único de estos buscadores es Exalead que de alguna manera si utiliza una estructura de metadatos creada por ellos, la cual cuenta con 12 campos en los cuales se indizan los datos de cada documento que se anexa a sus índices, lo que permite que se pueda recuperar la información adecuadamente, otro aspecto importante de este buscador es que cuenta con el apoyo de clusters dinámicos y ontologías, que hacen más accesible la búsqueda de información y al utilizar metadatos todos los documentos que recupera, pueden ser consultados en su totalidad.

Conclusiones

La evaluación sobre el funcionamiento de diferentes herramientas de búsqueda y recuperación de la información que se encuentran en la Web, así como la identificación de si existe relación entre la recuperación de la información y el uso de metadatos permitió confirmar los supuestos planteados al inicio del trabajo, en donde se menciona que, si se utiliza una estructura de metadatos adecuada en la organización de la información, se obtiene un mayor índice de recuperación. Sin embargo, muchas de las herramientas de búsqueda de información de la web no cuentan con estructuras de metadatos adecuadas, lo que dificulta la recuperación de la información, en particular la recuperación temática.

En tal sentido, como aspecto de alta importancia se menciona el incremento de información en Internet, es una muestra del desarrollo que han alcanzado las tecnologías de información, al crear herramientas que le permiten a los usuarios de la red poder buscar y recuperar información para sus investigaciones, además de permitirles un espacio en sus servidores para que puedan subir sus documentos de interés a la Web, ocasionando que se incremente el número de información día con día, sin ningún elemento para su organización.

Asimismo, existe un gran porcentaje de usuarios de la red, que representan un segmento de alta importancia por la función que desempeñan e identificados como los profesionales de la información, quienes contribuyen de forma asertiva en los procesos de búsqueda y recuperación de información. En líneas generales, poder superar las debilidades que surgen al momento de recuperar la información por causas de las herramientas de búsqueda, llámese motores de búsqueda, metabuscadores y buscadores semánticos, requiere de la identificación de una serie de características, las cuales al ser resueltas facilitan la recuperar la información, siempre y cuando el usuario sea capaz de identificarlas y las solventa adecuadamente.

En tal sentido, las herramientas que han sido analizadas, cuentan con diferentes formas de realizar búsquedas en sus índices para recuperar información, por lo que se requiere claridad sobre la manera de presentar la información y de no ser correcta, falló la forma de utilizar las estructuras de metadatos adecuadas. Es por

ello, necesario que al momento de realizar búsquedas de información se recuperen una gran cantidad de documentos, haciendo muy difícil para el usuario seleccionar los de mayor pertinencia entre tantos.

De tal manera que los motores de búsqueda convencionales, son y seguirán siendo una puerta para acceder a la información que se encuentra en la red, pero desafortunadamente tienen limitantes que no les permiten recuperar información relevante que se encuentran en algunos sitios de Internet. Con respecto, a los metabuscadores, al utilizar las plataformas de los motores de búsqueda, tienden a recuperar menos documentos, pero también sufren diferentes problemas en la recuperación, debido a la falta de bases de datos.

Finalmente en el caso de los buscadores semánticos, solamente el buscador Exalead, fue el único que recupero información de calidad y en la presentación de resultados también fue el único que permitió visualizar todos los documentos que recupero. Los resultados de la investigación presentada, conllevan a establecer las bases para realizar otros que complementen el supuesto de la existencia de una relación entre el uso de estructuras de metadatos y la recuperación de información eficiente.

Referencias

- BlogPulse: una herramienta de medición de Nielsen (2010). Recuperado el 11 de noviembre del 2016, de <http://www.blogpulse.com>
- Caplan, P. (1995). You call it corn, we call it syntax-independent metadata for documentlike objects. *The Public Access Computer Systems Review*, 4(6). Recuperado el 18 de diciembre del 2016, de <http://epress.lib.uh.edu/pr/v6/n4/capl6n4.html>
- Data Never Sleeps How much Data is Generated Every Minute. (2012). Recuperado el 18 de diciembre del 2016, de <http://www.domo.com/blog/2012/06/how-much-data-is-created-every-minute/>
- eBizMBA (2016) Top 15 Most Popular Search Engines. Recuperado el 4 de diciembre del 2016, de <http://www.ebizmba.com/articles/search-engines>
- Escobar, L., Paredes, A., y Barroso, M. (2015). Competencias del profesional de la información ante la evolución de las fuentes de información digitales. *Accesbib: Revista de Bibliotecología y Ciencias de la Información*, 4(1), 53-66. Recuperado de <http://www.revista.accesbib.org/index.php/revacc/article/view/3/4>
- El universo digital: ¿Cuánta información existe en Internet? (2010) .Recuperado el 1 de diciembre del 2016, de <http://w5.cl/2011/09/14/el-universo-digital-%C2%BFcuanta-informacion-existe-en-internet/>
- Gantz, J. y Reinsel, D. (2011) Extracting Value from Chaos. Recuperado el 8 de diciembre del 2016, de http://www.emc.com/digital_universe.
- Gantz, J. y Reinsel, D. (2010) The Digital Universe Decade-Are You Ready. Recuperado el 18 de diciembre de http://www.emc.com/digital_universe.

- Levene, Mark (2010) An introduction to search engines and Web navigation. Recuperado el 18 enero del 2017, de <https://luwj.googlecode.com/svn/trunk/Ebook/An%20Introduction%20to%20Search%20Engines%20and%20Web%20Navigation%202nd.pdf>
- Listofsearchengnes (2016) Top 10 Search Engines List. Recuperado el 18 de diciembre de <http://www.listofsearchengines.org/>
- Pinto, M. y Lancaster, F. (2001). “Hacia la modelización del documento científico” en Procesamiento de la información científica. Madrid: Arco/libros, p. 41-67
- Qué es un buscador. (2016). Recuperado el 18 de diciembre de <http://www.quees.info/que-es-un-buscador.html>
- Reliablesoft (2016) Top 10 search engines in the World. Recuperado el 18 de diciembre del 2016, de <https://www.reliablesoft.net/top-10-search-engines-in-the-world/>
- RDF Working Group, Resource Description Framework (RDF), (2014). Recuperado el 20 de noviembre del 2016 de <http://www.w3.org/RDF/>
- Top 7 Semantic Search Engines as an Alternative to Google. (2016) [en línea] Recuperado el 5 de enero del 2017, de <http://www.makeuseof.com/tag/top-7-semantic-search-engines-alternative-google-search/>
- 9 Semantic Search Engines That Will Change the World of Search.(2016). Recuperado el 18 de diciembre del 2017, de <https://www.searchenginejournal.com/semantic-search-engines/9832/>
- WWW.quees.info. Qué es un buscador. (2016). Recuperado el 5 de enero del 2017, de <http://www.quees.info/que-es-un-buscador.html>