Universidad Nacional de San Luis
Facultad de Ciencias Humanas
Licenciatura en Comunicación Social – Licenciatura en Periodismo Licenciatura en Producción de Radio y Televisión

Material Teórico (Investigación en Internet- Parte I)

Tecnología de la Comunicación I

Prof. Resp. Esp. Viviana Mercedes Ponce
Prof. Aux. Esp. Ana Silvina Maldonado
Prof. Aux. Lic. Pablo Oro

Elaborado por la Prof. Viviana Mercedes Ponce (vmponce@unsl.edu.ar)
Creative Commons Reconocimiento-Nocomercial-compartirigual 3.0 Unported License

2º Cuatrimestre – 2014

“Investigación en Internet y Búsquedas en la WWW: Visible, Invisible y Semántica”

Prof. Viviana Ponce

vmponce@unsl.edu.ar



1- Introducción- Búsquedas de información

Desde sus orígenes, y especialmente en los últimos años ha cobrado especial importancia el proceso de realización de búsquedas en Internet[1]. ¿Por qué?

Sencillamente porque la red de redes provee acceso a una gran cantidad de información diseminada en miles de millones de servidores web[2] y ésta, se constituye en un recurso muy valioso para todas las personas, no importa su profesión, su condición social, sus intereses, etc.

---

[1] *Internet:* red de redes de computadoras interconectadas entre sí, a través de medios físicos y protocolos de comunicación. Los medios físicos son aquellos que se componen de cables, conectores de red, dispositivos de interconexión (hubs, switch, router, etc.); mientras que los protocolos de comunicación (TCP/IP) son los que les permiten a las computadoras “hablar” e “interactuar” entre ellas.

Internet funciona a través de la lógica cliente-servidor: el servidor es una computadora de la red que contiene la información a consultar y el cliente es la computadora del usuario desde la que se busca información. El cliente realiza los requerimientos y el servidor es quien responde a los mismos, siempre que pueda hacerlo.

[2] De acuerdo al informe *The World Factbook* (https://www.cia.gov/library/publications/the-world-factbook) elaborado por la CIA (Central Intelligence Agency) en el que se reúnen datos de 267 países del mundo: históricos, gubernamentales, económicos, etc.; el país con mayor cantidad de servidores en el mundo es EEUU, con 505.000.00 (no indica año) y le siguen Japón con 64.453.000 (2012) y Brasil con 26.577.000 (2012). Argentina ocupa el puesto N° 13, con 11.232.000 de servidores (2012).

En Internet puede encontrarse simplemente de todo, pero a nosotros nos interesa aquel "todo" que teniendo un valor real, útil y verdadero nos permita no solo aprender e informarnos, sino también crear información nueva.

El proceso de búsqueda de información en la WWW[3] no es un acto sencillo, requiere de un método exhaustivo y organizado. La WWW es una telaraña compleja, y no todos los sitios se encuentran ajustados a las distintas necesidades de los usuarios. Se dice que Internet es variada, y en este sentido, la WWW no está atomizada, sino que se encuentra clasificada: la Web Visible, Invisible y Semántica, junto con los buscadores que proveen en cada caso, responden a los diferentes requerimientos de los usuarios que son los que, en definitiva, demandan el acceso a la información. Además, la información, en muchos casos debe ser evaluada puesto que como todos sabemos, Internet también es "libre[4]" y "autoregulada[5]" y en consecuencia, no existe un control de veracidad total sobre los contenidos que a través de ella se publican.

Considerando entonces las implicaciones anteriores, es posible definir a la investigación en Internet como: "*un proceso exhaustivo y organizado, a partir del cual es posible buscar información en la WWW (Visible, Invisible y Semántica)[6], interpretarla y discriminarla de acuerdo al análisis de sus resultados*".

En los siguientes apartados exploraremos un poco más este proceso, atendiendo a la definición y explicación de algunos conceptos involucrados.

2- Desmitificando conceptos equivocados en torno a las búsquedas

Algunas veces cuando buscamos información partimos de algunos supuestos previos que son considerados erróneos, puesto que no tienen una implicación verdadera a la hora de efectuar búsquedas. Por este motivo, es importante recordar algunas cuestiones que inciden en las mismas y que nos afectan, a saber:

---

[3] *WWW:* se conoce así a la telaraña de alcance mundial formada por una gran cantidad de páginas web enlazadas. Para acceder a las mismas es necesario conocer su dirección virtual o más específicamente, su dirección URL (Uniform Resource Locator, en español Localizador Uniforme de Recursos). La dirección URL de una página debe ser ingresada a un navegador de páginas web, como Internet Explorer, Google Chrome, Mozilla, etc.), para poder ser visualizada. Un ejemplo de dirección URL es la del Blog de la Cátedra: http://tecnologia2sl.blogspot.com

[4] Al decir que Internet es "libre" estamos diciendo que los usuarios tienen la libertad de expresarse, acceder y publicar información, comentarla, compartirla, etc. El problema es que en muchos casos la información publicada no es controlada.

[5] La autorregulación hace referencia a la capacidad de una organización o entidad para regularse a sí misma, en base al control y monitoreo voluntario.

[6] La WWW se encuentra clasificada en Visible, Invisible y Semántica. Más adelante, iremos profundizando en la definición y características de cada una de ellas.

- *Internet no es una biblioteca:* La red de redes de computadoras ofrece múltiples posibilidades y el libre acceso a la información es una de ellas. Los miles de millones de sitios web[7] accesibles desde la navegación por la WWW se encuentran dispersos en los miles de millones de servidores que los almacenan. En este sentido, los buscadores de información pueden contribuir a la organización o catalogación de los sitios y de sus recursos: textos, imágenes, audios, videos, archivos en diferentes formatos, etc., con el objeto de que puedan ser encontrados por el usuario a través de su correcto y efectivo uso.

- *Toda la información que está en Internet es cierta y vale:* al momento de publicar contenidos en la WWW no existe un control global que permita diferenciar la información verdadera de la que no lo es. Al ser libre, Internet permite que se publique información que en muchos casos no es controlada, por lo que circulan contenidos falsos. Los buscadores de información pueden proveernos acceso a datos reales y verdaderos, pero esta no es una característica que se plasme en todos ellos. Internet es "autoregulada", esto es, pueden existir controles al interior de los sitios, o bien, controles al interior de un país o países determinados, pero no existen controles a nivel mundial.

- *En Internet se encuentra toda la información:* la red de redes crece día a día a pasos agigantados. Miles de sitios se crean cada un período corto de tiempo y aunque se eliminen otros (la cantidad de sitios disponibles en la WWW es un número dinámico, en constante cambio) la cantidad de información disponible va exponencialmente en aumento. Asimismo esta realidad no condiciona el hecho de que podamos encontrar todo lo que buscamos: en la WWW solo podemos encontrar lo que alguien o algo[8] alguna vez publicó, y cabe pensar además, en si la información publicada posee un valor real, útil y verdadero. Los buscadores pueden brindarnos respuestas a lo que investigamos pero estas pueden no ser las únicas, en cuyo caso, puede ser necesario recurrir a otras fuentes de acceso a la información no digitales (gobierno, instituciones, organizaciones sociales, etc.).

---

[7] De acuerdo al Informe de *NETCRAFT* (http://news.netcraft.com/archives/2013/05/03/may-2013-web-server-survey.html), una compañía de servicios de Internet con sede en Bath, Inglaterra, existen al mes de mayo del presente año 672,837,096 sitios, 23.8 millones más de los que habían en el mes de abril.

[8] La información puede ser publicada no solo por personas, sino también por "robots" que en algunos casos realizan rastreos de la información de los principales medios de comunicación online y las presentan al usuario a través de un sitio web. Como ejemplo se puede mencionar a Google Noticias, que se define a sí mismo como un "agregador y buscador de noticias".

- *Hay información sobre todos los temas y sobre todo el mundo:* en la WWW no se encuentra toda la información, ni esta se refiere tampoco a todos los temas y acontecimientos ocurridos en los diversos lugares del mundo. Lo que sí es necesario considerar es que existen países (como EEUU por ej.) productores y difusores de gran cantidad de información que en muchos casos debería ser considerada, sobre todo cuando nos encontramos realizando una investigación profunda en torno a un tema determinado. Independientemente de las barreras que se le podrían presentar al investigador por el desconocimiento de algún idioma, lo cierto es que, afortunadamente, algunos buscadores de información como por ej. Google[9] proveen traductores interesantes, que ayudan y facilitan la lectura en español de sitios extranjeros.

- *Plagiar es ingenioso porque todo es público:* la falta de control que muchas veces no se aplica a los contenidos[10], hace que en muchos casos se publique información sin tener la autorización de su/s autor/es, o lo que es peor, que se publique información acreditándose la autoría como propia. Lo que se encuentra en Internet no puede reutilizarse así nomas, es necesario dar crédito a sus productores y respetar las restricciones que éstos han indicado para sus obras (literarias, artísticas y digitales). Cobran sentido entonces, los conceptos de "propiedad intelectual de una obra", que requiere "*respetar todo conocimiento producido por el intelecto humano que se puede plasmar en un objeto tangible o intangible*" y los "derechos de autor", que especifican las "*normas que hacen a la distribución, producción y reproducción de sus contenidos*". La producción de información nueva a partir de otra ya existente, requiere estar atento entre otros, a los conceptos de "Copyright[11]" y "Copyleft[12]" relacionados con la creación de una obra. Algunos buscadores de información proveen herramientas que nos permiten encontrar contenidos que se pueden usar, compartir, o modificar libremente, entre otros.

---

9 *Google:* empresa multinacional estadounidense especializada en productos y servicios relacionados con Internet, software, dispositivos electrónicos y otras tecnologías.

10 Esta situación ha sido y es aún motivo de debate en innumerables países del mundo que buscan proteger los derechos de autor de sus principales artistas y literarios. El siguiente video: http://www.youtube.com/watch?v=oPrCS2PmWyo nos invita a pensar en algunos de estos aspectos.

11 *Copyright:* conjunto de leyes, sancionadas por cada país, que le otorgan al creador de una obra los derechos exclusivos sobre la misma durante un tiempo limitado. Generalmente este derecho tiene que ver con el "derecho a copia" pero también se relaciona con el derecho a ser reconocido por la creación de la obra, a ser acreditado para el trabajo, a la determinación de quién puede adaptarlo a otras formas, quién puede beneficiarse económicamente de ella, y otros derechos relacionados.

12 *Copyleft:* es la práctica de utilizar la ley de derechos de autor para ofrecer el derecho de uso, distribución de copias y modificación de una obra, siempre que se reconozca y preserve al autor original de la misma.

Luego de estas consideraciones y antes de comenzar a efectuar las búsquedas, ya estamos en condiciones de introducirnos a las herramientas informáticas que nos permitirán desarrollarlas: los "buscadores de información", que se describen en el siguiente apartado.

3- Buscadores de información: definición, clasificación, características y ejemplos

Como lo expresa Jaramillo Marín (2012)[13] la búsqueda en Internet es una de las herramientas más útiles en la vida de un comunicador, y especialmente un periodista. La posibilidad de acceder a la información disponible de forma ordenada, rápida y relevante facilita el trabajo de quienes tienen por objetivo informar al público sobre lo que ocurre en el mundo. Pero la búsqueda en Internet ha evolucionado y ya no es aquel proceso al que estamos acostumbrados. Hoy existe mucha información para organizar y muchos más formatos y plataformas para extraer información en tiempo real.

Los buscadores de información pueden definirse como: "*herramientas informáticas especializadas que nos posibilitan encontrar en la WWW diversidad de tipologías de información digital: textos, audios, videos, animaciones, mapas, archivos en diversos formatos, etc."*

Sus diversas características hacen que sean clasificados de diferentes formas según diferentes autores. En esta asignatura, tendremos en cuenta la siguiente tipología de buscadores, puesto que a partir de las mismas es posible encuadrar al resto de los tipos de buscadores emergentes en los últimos años.

- Buscadores por temas o categorías
- Motores de búsqueda
- Metabuscadores y Multibuscadores

A continuación se describe cada una de ellas.

Buscadores por tema o categorías

Este tipo de buscadores, denominados también "por directorio" construyen sus bases de datos en base a la información proporcionada por personas o robots,

[13] Marín Jaramillo M. (2012). "*Guía para periodistas*". Bogotá, Colombia. Disponible en: http://www.slideshare.net/EDWINLOZADA/google-para-periodistas

que rastrean la WWW en busca de información y la organizan en temas o categorías.

Con la premisa de proporcionarle al usuario información confiable, en el primer caso, las bases de datos se alimentan de los aportes de usuarios expertos de la red que evalúan los sitios para determinar si se incluyen o no en el directorio[14] mientras que en el segundo, se trata de programas de computadora que si bien no tienen la capacidad de evaluar sitios, arman sus bases con aquellos más destacados en cuanto a la información que proveen. La cantidad de información recopilada por un robot supera a la provista por un humano, no solo por la cantidad de programas dedicados al rastreo, sino también por el tiempo que le destinan a la realización de este proceso.

Los buscadores por directorio se encuentran organizados de acuerdo a una estructura jerárquica en la que la información se concentra desde lo más general a lo más específico, estableciendo con esto, un orden en la realización de la búsqueda mediante la navegación por los diferentes temas o categorías. Además, permiten efectuar búsquedas a través de una palabra clave que permite obtener resultados sin necesidad de saber de antemano la categoría en la que estos se pudieran haber encontrado. Sus resultados son devueltos en lista ordenados alfabéticamente, o por nivel de popularidad (Page Rank)[15].

Pueden ser generalistas (es decir, proveer enlaces a sitios web basados en una amplia diversidad de temas) o especializados (brindan enlaces que se abocan a una temática en particular).

Como ejemplos de este tipo de buscadores se pueden mencionar:

- *a nivel mundial y generalista:* DMOZ (http://www.dmoz.es/) un proyecto de directorio abierto editado por una gran comunidad de editores voluntarios que se rigen por la filosofía del código abierto, esto es, no hay y nunca habrá un costo por sugerir o usar datos del mismo.
- *a nivel mundial y especializado*: Google Noticias (https://news.google.com.ar/) un programa automatizado que rastrea de forma constante la información de los principales medios de comunicación online y Technorati Blogs (http://technorati.com/blogs/directory/) que permitía hasta hace un tiempo (ya que en la actualidad no se encuentra disponible) la realización de búsquedas de blogs en inglés, en su gran cantidad de bases de datos.

---

[14] En algunos casos estos buscadores permiten que el usuario común pueda sugerir un sitio web, el cual será posteriormente revisado por los editores expertos antes de ser incluido en el directorio.

[15] *PageRank* (PR): valor numérico que representa la importancia de una página web en Internet. Se mide de acuerdo a cada enlace (link o voto) que se coloque a una página.

- *a nivel regional y especializado:* se puede mencionar a abcmedico.com (http://www.abcmedico.com/), un directorio médico de España.

Motores de búsqueda o Buscadores por palabra clave

Construyen sus bases de datos en base a un rastreo de la WWW realizado de forma periódica por robots[16] que se conoce como "spidering" y almacenan no solo páginas web sino también otro tipo de archivos: .doc, .pdf, .jpg, .gif, etc.

No se encuentran organizados por medio de una estructura jerárquica, sino que disponen de la información a nivel general, no clasificada. La manera de efectuar búsquedas por parte del usuario es a través de "palabras clave" que permiten especificar una o más palabras relacionadas con el/los conceptos que se está/n buscando.

Sus resultados pueden ser devueltos de diferentes maneras, ya sea, mediante el concepto de PageRank (esto es, las páginas más linkeadas o votadas serán las que aparezcan como parte de los primeros resultados de una búsqueda) por fecha (ordenados de acuerdo a uno o más periodos de año determinados), por país o por idioma.

Pueden ser generalistas o especializados y al estar provistos por una gran cantidad de operaciones suelen brindar manuales de ayuda que posibilitan aprovecharlos al máximo.

Por otra parte, y a diferencia de lo que ocurría en el caso de los buscadores por directorio, los robots que arman las bases de datos recorren toda la WWW sin discriminar entre sitios confiables o no, por lo que en muchos casos pueden devolver información falsa como resultado. En este caso, se deberá realizar una evaluación de la información antes de considerarla como verdadera.

Como ejemplos de este tipo de buscadores se pueden mencionar:

- *a nivel mundial y generalista:* Google (http://www.google.com.ar) considerado como el buscador más diverso y potente en la actualidad, Bing (http://www.bing.com/?setlang=es) y Google Maps (http://maps.google.com.ar) un localizador geográfico de imágenes vía satélite de diversos lugares del mundo.
- *a nivel mundial y especializado:* Google Imágenes (http://images.google.es/), Google Videos (http://www.google.es/videohp) para la búsqueda de imágenes y videos respectivamente, Google Books (http://books.google.es/) para libros, Google Blogs (https://www.google.com.ar/blogsearch?gws_rd=ssl) para blogs y Google

[16] También conocidos como spiders, arañas, bots o indexadores.

Scholar (http://scholar.google.es/) para las consultas de información de tipo académica. Los buscadores de personas Socialmention (http://www.socialmention.com/) y Twitterfall (http://twitterfall.com/) también pueden considerarse dentro de esta categoría.

Buscadores de buscadores

Realizan búsquedas de información en uno o más buscadores de información existentes, por lo que no poseen una base de datos propia.

Pueden ser generalistas o especializados y permiten especificar las búsquedas mediante algunos operadores de búsqueda que proveen.

Son algo lentos y pueden devolver resultados repetidos.

Se clasifican en:

- *Metabuscadores:* muestran los resultados en una única ventana, indicando para cada uno de ellos el buscador en el que se encontró la información. Como ejemplos se puede mencionar a Ixquick (https://www.ixquick.com/esp/), un metabuscador con sede en Nueva York y Holanda que permite la realización de búsquedas no solo en otros buscadores, sino también en su propio directorio de imágenes, videos y telefónico y Dogpile (http://www.dogpile.com/), que permite encontrar información en los mismos buscadores que Metacrawler y al igual que este último permite acceder a sus páginas blancas y amarillas.
- Multibuscadores: muestran los resultados en distintas ventanas, una por cada buscador. Como ejemplo se puede mencionar a The Info (http://www.theinfo.com/) un multibuscador que si bien, no ha sido actualizado desde el año 2000, permite la realización de búsquedas de películas, música y noticias, entre otros.

En la actualidad existen algunas tendencias que apuntan a un mayor uso de los buscadores de tipo motores de búsqueda, por sobre los buscadores de tipo directorio o de tipo meta/multibuscadores. El siguiente cuadro comparativo resume sus principales aspectos:

	DIRECTORIOS O CATEGORÍAS	**MOTORES DE BÚSQUEDA**	**BUSCADORES DE BUSCADORES**
¿Cómo se construye la	A través de un grupo humano o robot	Por medio de un robot	Por medio de un robot

información?			
¿Dónde buscan la información?	En la WWW	En la WWW	En motores de búsqueda existentes
¿Qué información buscan?	General y Especializada	General, Especializada y Específica	General, Especializada y Específica
¿Cómo devuelven sus resultados?	Ordenados por orden alfabético o Page Rank	Ordenados por Page Rank	Ordenados por Page Rank
¿Cuántos resultados devuelven?	Pocos o Muchos	Muchos	Muchos
¿La información devuelta es correcta?	Sí, en algunos casos puede estar desactualizada	No en todos los casos, hay que evaluarla	No en todos los casos, hay que evaluarla
¿Cómo se busca?	A través de las categorías o ingresando una palabra clave	A través de una palabra clave	A través de una palabra clave
Facilidad de uso	Sencilla	Compleja	Media

4- La WWW Visible: definición y características

La WWW "Visible" o "Superficial" puede definirse como: *"…la porción de la Web que es vista y consecuentemente indexada*[17] *por los buscadores comunes*".

Esta porción de la Web se compone de imágenes, páginas web (en formato .HTM/HTML y .PHP) y archivos: de texto, presentaciones con diapositivas, hojas de cálculo, ejecutables, portables, de animaciones y de audio y video, entre otros[18]. El resto de la información, como veremos más adelante, no podrá ser

[17] *Indexar:* Acción de registrar ordenadamente datos e informaciones para elaborar su índice (diccionario de la Real Academia Española: http://www.rae.es/rae.html).

[18] En el estudio *How much information*? realizado en el año 2003 por un grupo de profesores y alumnos de la Escuela de Gestión y Sistemas de Información en la Universidad de California en Berkeley (http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/internet.htm) cuyo objetivo es

encontrada por diversas razones, que en algunos casos tienen que ver con limitaciones en el funcionamiento interno de cada buscador, y en otros, con restricciones impuestas desde los propios sitios web, que no desean, no pueden o no deben ser visualizados.

Entre sus principales características se tiene:

- *Está compuesta por un conjunto de páginas estáticas públicamente accesibles*, es decir, por archivos en formato .HTM/HTML y en algunos casos .PHP, que contienen información permanente, que no cambia en tiempo real, y que no posibilita la interacción con el usuario (ya sea a través de foros, chat, comentarios, etc.), como cuestiones principales[19].
- *Constituye una porción relativamente muy pequeña de toda la Web*, un 18% aproximadamente, de acuerdo a un último estudio realizado en el año 2003, por profesores y alumnos de la Universidad de Berkeley, en California, EEUU[20].
- *La información que se recupera debe ser cuidadosamente evaluada y validada*, puesto que en algunos casos puede ser obtenida por robots (como Googlebot[21] en el caso del motor de búsqueda Google) que trabajan con software informático que rastrea la WWW, pero que al tratarse de un programa de computadora, no tiene la capacidad para discriminar o evaluar la información.

Entre las principales herramientas informáticas que se pueden utilizar para encontrar información en esta parte de la WWW, se tienen los buscadores por directorio, los motores de búsqueda y los buscadores de buscadores comunes que ya han sido descriptos en los apartados anteriores. Estos nos permiten la realización de operaciones básicas y avanzadas con las que es posible acceder a la diversidad de los tipos de archivos que la componen.

---

determinar la cantidad nueva de información que se crea cada año, se puede apreciar concretamente cuál es el porcentaje de archivos de diverso tipo existentes hasta esa fecha.

[19] En el sitio de *Art Factory* (http://www.artfactory.es/noticias/detalle/noticia/00032/web-estatica-vs-web-dinamica/) dedicado al diseño gráfico, diseño web y comunicación corporativa para la empresa, se puede leer un artículo interesante y completo, respecto de las características y diferencias entre páginas web estáticas y dinámicas.

[20] En el estudio How much information? (http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/internet.htm) puede verse que existen 167 terabytes en la Web Visible contra 91850 terabytes que conforman la Web Invisible.

[21] GoogleBot: robot de rastreo web de Google, que descubre páginas nuevas y actualizadas y las añade a su índice : https://support.google.com/webmasters/answer/182072?hl=e

5- Buscadores por directorio, temas o categorías

Conforman una valiosa base de datos de información revisada y validada por personas o robots, que generalmente es confiable y que puede no resultar suficiente en cuanto a su cantidad y actualidad.

¿Cómo buscar?

El usuario ejecuta las búsquedas de dos formas posibles, dependiendo de si conoce o no, el tema o la categoría en la que podría encontrarse la información que está buscando. Las estrategias que puede utilizar en este tipo de buscadores son[22]:

1) Si conoce el tema, identifica la categoría en la que este se ubica y continúa con el paso 2.a). Si no conoce el tema realiza el paso 2.b).

2) Realiza alguno de los siguientes pasos:

2.a) Busca la categoría en las categorías existentes. Se dirige al paso 3.

2.b) Escribe la palabra clave relacionada con lo que está buscando. Se dirige al paso 3.

3) De los resultados devueltos observa si se encuentra lo que busca.

4) Registra luego el/los resultados obtenidos.

¿Cómo participar?

Algunos buscadores temáticos permiten que el usuario participe de sus bases de datos mediante la realización de aportes o sugerencias de sitios web que pueden resultarles de interés a otros usuarios. Existen dos categorías posibles:

- El usuario común[23], que sugiere simplemente un sitio que será puesto después a consideración por parte del grupo de expertos que mantiene la base de datos del directorio.

[22] En la presentación con diapositivas correspondiente a la Clase 2 de esta asignatura puede consultar ejemplos donde se realizan búsquedas en DMOZ y Google Noticias, cuando se conoce o no, el tema o categoría en el que la información se encuentra.

[23] El usuario común puede sugerir su propio sitio web si desea que este aparezca en las bases de datos del directorio. Para ello, puede utilizar por ej. el enlace "Agregar URL" en el buscador DMOZ: http://www.dmoz.org/help/submit.html que le informará entre otros aspectos, cómo agregar un sitio, cuánto tiempo tarda el mismo en ser visualizado y cómo hacer para que aparezca en más de una categoría, entre otros.

- El usuario editor[24], considerado un experto en una o más temáticas y con la autoridad suficiente para agregar, revisar y/o eliminar sitios web de las bases de datos.

Otras posibilidades que ofrecen

Algunos buscadores por directorio brindan la posibilidad de realizar operaciones que al usuario le puede resultar interesantes, como por ejemplo:

- *Brindar información acerca de los temas o categorías:* en cuanto a la información que contienen y a la/s posible/s categorías en las que se podría encontrar más sitios relacionados.
- *Seleccionar el parámetro por el que serán devueltos los resultados de una búsqueda:* esto es, por fecha, por orden alfabético o por orden de importancia, entre otros.
- *Personalizar los temas que consulta con más frecuencia, como así también las fuentes de donde proviene la información:* en el caso de Google Noticias (http://news.google.es/) el usuario puede realizar ambas acciones, dependiendo del tipo de información en el que éste se encuentre abocado o interesado y los medios de comunicación que siga, o que considere a su criterio, más confiables.

6- Motores de búsqueda o Buscadores por palabra clave

Generalmente este tipo de buscadores constituye la opción más potente para la exploración de la Web Visible principalmente porque disponen de una gran cantidad de operadores que permiten especificar las búsquedas y porque devuelven una gran cantidad de resultados que pueden constituirse en una poderosa fuente de información.

¿Cómo buscar?

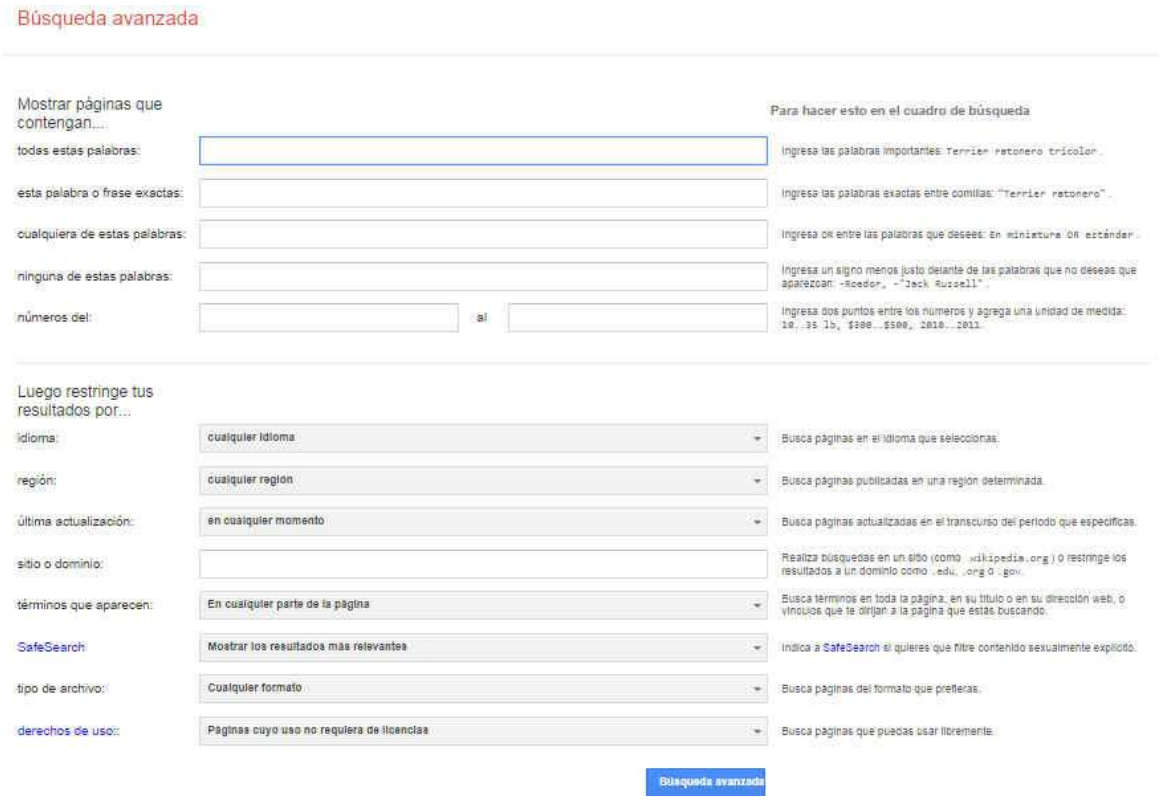
[24] El buscador DMOZ permite que nuevos usuarios con los debidos conocimientos realicen aportes a su directorio bajo la filosofía del software libre y abierto, esto es, el software puede compartirse, reutilizarse y modificarse libremente, conservando las licencias de su autor original. En el enlace "Hazte editor" del sitio de DMOZ: http://www.dmoz.org/cgi-bin/apply.cgi?where=/ se puede consultar acerca de qué es un editor, cuáles son los requerimientos que deben cumplirse para serlo, en qué categorías puede participar y cuál es la compensación que éste podría obtener por su trabajo.

El usuario ejecuta las búsquedas a través de la introducción de una “palabra clave” (compuesta por una o más de una palabra) que resume, sintetiza o especifica aquello que se está buscando y a la que se le puede pedir que reúna una o más condiciones. Las estrategias que utiliza el usuario en este tipo de buscadores son:

1) Coloco la «palabra clave» en la caja de búsqueda.

2) Selecciono una o más opciones de la caja de búsqueda avanzada

3) Cliqueo en el comando “Búsqueda Avanzada”.



En la clase de teoría (http://www.slideshare.net/vmponce/clase-2-investigacin-en-internet-38808857) desde la diapositiva 34 hasta la 60 aproximadamente, puedes sobre la diversidad de formas de búsqueda que puedes realizar, haciendo uso del cuadro anterior de búsqueda.

No obstante, es importante que tengas en cuenta además que una vez que se es un usuario avanzado en las búsquedas de información, estas se pueden realizar directamente haciendo uso de los operadores de búsqueda. El siguiente apartado

te explica cómo puedes realizar este proceso de una forma más avanzada. Si igualmente te resulta complicado, puedes guiarte directamente con las diapositivas que se mencionaron anteriormente.

¿Cómo especificar la búsqueda mediante operadores o parámetros de búsqueda?

Los motores de búsqueda proveen diversidad de operadores que se pueden utilizar en la construcción de la palabra clave. El uso adecuado de los mismos nos permite encontrar rápidamente lo que deseamos, de forma eficaz y eficiente.

A la clásica y tradicional búsqueda en la que no se utilizan operadores, se opone aquella que sí hace de los mismos, con resultados mucho menos extensos y más adecuados. A continuación se describen las distintas formas de buscar[25] en el motor de búsqueda Google (http://www.google.com) el más utilizado a nivel mundial.

- *Operador “”*

Se utiliza cuando se desea buscar una palabra o frase exacta. Su uso tiene sentido cuando al escribir una palabra clave, sin ningún tipo de operador, por ej.

Periodismo digital en Argentina

…nos encontramos con que nos aparecen resultados sin sentido y que no son relevantes para lo que deseamos encontrar, como por ej.

Resultado 1:
Cómo trabajan los medios digitales en la Argentina / Inicio / FOPEA ...

Resultado 2:
La Asociación de Editores Digitales de la Argentina tiene por misión ...

Resultado 3:
Google premia la innovación periodística en la era digital - Google+, entre otros.

Además, es importante considerar el modo en el que Google devuelve los resultados. Al ser un motor de búsqueda inteligente considera todos aquellos

[25] En la presentación con diapositivas correspondiente a la Clase 2 de esta asignatura puede consultar ejemplos donde se realizan distintos tipos de búsquedas con distintos operadores y parámetros en el motor de búsqueda de Google.

términos que se ajustan o asemejan a la palabra clave y que aparecen además en distintos lugares, ya sea en la dirección URL, como en el contenido o título de la página (en el resultado 1 periodismo se reemplazó por medios y digital por digitales). ¡Realice esta búsqueda y verifique estas acciones!

Entonces, y volviendo al uso del operador doble comilla, cabe mencionar que tiene sentido de ser utilizado cuando deseamos que el buscador nos devuelva los resultados en ese orden y con esa frase exacta, esto es, al ingresar la palabra clave encerrada entre dobles comillas:

"Periodismo digital en Argentina"

…encontraremos resultados donde aparezca toda esta frase y en el orden propuesto:

Resultado 1:
Investigando sobre **periodismo digital en Argentina**…

Resultado 2:
Periodismo Digital en Argentina - Scribd…

Resultado 3:
Libro: "**Periodismo Digital en Argentina**" | Sicrono…, entre otros.

El uso del operador "" reduce la cantidad de resultados y nos devuelve solamente aquellos que se ajustan a la búsqueda concreta que deseamos realizar. No siempre es conveniente utilizarlo, ya que en algunos casos el buscador puede no devolvernos ningún resultado.

- *Operador OR*

Busca una o más palabras claves a la vez. Un ejemplo de este tipo de búsqueda es: **"periodismo digital" OR "periodismo ciudadano"**

…que nos devolverá como resultados páginas o archivos que contengan "periodismo digital" o "periodismo ciudadano" o ambos.

- *Operador AND*

Busca todas las palabras claves a la vez. Por ej. podemos escribir:

"periodismo digital" AND "periodismo ciudadano"

…que devolverá como resultados páginas que contengan “periodismo digital” y “periodismo ciudadano” al mismo tiempo.

- *Operador + (Inclusor)*

Permite incluir una o más palabras clave que por alguna razón no son consideradas por el buscador como parte de sus resultados, como ocurre por ej., en el caso de las “StopWord[26]”. Un ejemplo de este tipo de búsqueda es:

Periódicos digitales +de Argentina …en la que forzamos que aparezca la preposición “de”.

- *Operador - (Exclusor)*

Permite excluir una o más palabras clave de los resultados de una búsqueda. Como ejemplo se puede mencionar: **Periódicos digitales -Argentina**

…que devuelve información de periódicos digitales que no pertenecen a la Argentina.

- *Operador * (Comodín)*

Busca cero o más palabras claves relacionadas a una búsqueda. Como ejemplo se puede mencionar: *** de Periodismo Digital**

…que puede devolver, entre otros, los siguientes sitios resultados:

Resultado 1:
Periodismo Digital - Comunicación 2.0 - Ramos Mejía - Diseño web ...

Resultado 2:
Apuntes de Periodismo Digital

Resultado 3:
Foro de Periodismo Digital Rosario

- *Operador filetype*

[26] *StopWord:* se les llama así a las palabras de significado vacío que no son registradas por las arañas de los buscadores a la hora de hacer una búsqueda. Están compuestas por artículos, pronombres y preposiciones.

Busca archivos con un formato determinado: pdf, ppt, doc, etc. Por ejemplo: **"periodismo digital" filetype:pdf** que devolverá archivos en ese formato que contengan la palabra periodismo digital.

- *Operador allintitle*[27]

Busca páginas que contengan la palabra clave en el título de la página. Por ejemplo: **allintitle: "radio digital"** nos devolverá como resultados un listado de páginas web cuyo tema central se relacione con la radio digital.

- *Operador: allintext*

Busca páginas que contengan la palabra clave en su contenido. Por ejemplo**: allintext: "radio digital"** nos devolverá como resultados un listado de páginas web donde se relacionen sus contenidos derivados con el concepto de radio digital.

- *Operador: allinurl*

Busca páginas que contengan la palabra clave en su dirección URL. Por ejemplo: **allinurl: "radio digital"** nos devolverá páginas web con la siguiente forma:

http: //www.**radio-digital**.net/…
http://mipagina.cl/**radio-digital**-fm.htm y …
http://www.softonic.com › TV y **Radio online** › **Radio online**…, entre otras.

- *Operador: site*[28]

Busca páginas que pertenezcan a un sitio o dominio determinado. Como ejemplos se puede mencionar:

"radio digital" site:.com que buscará páginas de tipo comercial relacionadas con la temática de la radio digital.

"radio digital" site:www.clarin.com que buscará páginas en el sitio web de Clarin.com que contengan o se relacionen con la radio digital.

- *Operador: related*

[27] *Allintitle:* operador de búsqueda denominado "title" o "intitle" en otros buscadores.
[28] *Site:* operador de búsqueda denominado "host" en otros buscadores.

Busca páginas cuyo contenido es similar a otra página. Por ejemplo, la búsqueda: **related: http://www.periodistadigital.com/** nos devolverá como resultado un listado de medios periodísticos en línea con información relacionada al periódico Periodista Digital.

- *Operador: link*

Busca páginas con enlaces a otra página. Por ejemplo: **link:** http://www.periodista**digital**.com/ nos devolverá como resultados un listado de páginas web que linkean o redireccionan al sitio web de Periodista Digital.

Otras posibilidades que ofrecen

Las operaciones que se pueden realizar a través de un motor de búsqueda son amplísimas, en relación a las que se podrían efectuar con un buscador por directorio. Entre las más importantes cabe mencionar:

- *Efectuar búsquedas por idioma, región o país, fecha o intervalo de fechas, seguridad (SafeSearch[29]) y derechos de uso de los contenidos que circulan por la Web:* al igual que otras acciones que se describieron anteriormente, estas pueden ser efectuadas desde el comando: "Búsqueda avanzada[30]", una opción muy utilizada a nivel mundial.
- *Efectuar búsquedas con operadores extras*: que permiten entre otros, determinar el estado del tiempo (por ej. **Tiempo San Luis**), definir un término (por ej. **Define: inherente**), realizar un cálculo matemático (por ej. **Calculadora (2 *5) + 3**), una conversión de moneda (por ej. **Conversor 100 dólares en euros**) y la localización de una ubicación geográfica (por ej. **Mapa San Luis**).
- *Efectuar búsquedas completas de tipologías concretas de archivos*: tales como imágenes, videos y libros, entre otros. En el caso de Google Imágenes (http://images.google.es/) es posible encontrar imágenes basadas en un cierto tamaño, color, tipología y fecha. En cuanto a Google Videos (http://www.google.es/videohp) se puede determinar su duración, fecha de producción, calidad y fuente de la que provienen. Con respecto a los libros (http://books.google.es/), se puede acceder a ebooks gratuitos, discriminar entre libros y revistas y ordenarlos de acuerdo a su importancia

[29] *SafeSearch:* filtro que permite cambiar la configuración del navegador de páginas web para evitar o reducir la aparición de contenido para adultos en los resultados de búsqueda.

[30] En el motor de búsqueda de la empresa Google esta acción se puede realizar desde el siguiente enlace: https://www.google.com.ar/advanced_search?

o fecha determinada. En todos estos casos particulares de búsqueda además, es posible acceder a la opción de "Busqueda avanzada" que en interacción con las búsquedas de información de Google nos brindará la posibilidad adicional de realizar búsquedas por idioma, región y derechos de uso, entre otras.

- *Configurar su funcionamiento:* en cuanto a la elección de la activación de las predicciones automáticas de palabras[31] o la búsqueda por voz, la selección de la cantidad de resultados por página y el lugar en el que estos serán devueltos y el manejo del historial de búsqueda.

7- Buscadores de buscadores

A la hora de devolver información son mucho más potentes que los motores de búsqueda por la sencilla razón de que concentran los resultados de dos o más buscadores. Sin embargo, y aunque permiten en muchos casos la realización de una búsqueda avanzada, a la hora de buscar, esta nunca llega a ser tan variada como la que proveen los motores de búsqueda más utilizados.

¿Cómo buscar en un meta buscador?

Las estrategias que utiliza el usuario en este tipo de buscadores son:

1) Coloca la «palabra clave» en la caja de búsqueda del meta buscador.

2) Ejecuta la búsqueda a través del botón de Búsqueda.

3) Examina los resultados y el buscador del cual provienen y registra uno o más de ellos si los considera adecuados.

¿Cómo buscar en un multi buscador?

Las estrategias que utiliza el usuario en este tipo de buscadores son:

1) Coloca la «palabra clave» en la caja de búsqueda del multi buscador.

2) Selecciona los buscadores en los cuales realizará la búsqueda.

2) Ejecuta la búsqueda a través del botón de Búsqueda.

3) Examina los resultados que son mostrados en distintas ventanas del navegador (una por cada buscador) y registra uno o más de ellos si los considera adecuados.

[31] Como Google Instant (http://www.google.es/instant/) por ejemplo, que muestra resultados en el buscador a medida se escribe.

¿Cómo especificar la búsqueda mediante operadores o parámetros de búsqueda?

Tanto los meta como los multi buscadores permiten que el usuario pueda especificar la búsqueda a través del uso de diversos operadores, que se pueden acceder mediante el comando de "Búsqueda Avanzada".

En el caso de Ixquick (http://www.ixquick.com) las búsquedas incluyen el uso del operador "", OR, -, host[32], en el título y en la dirección URL, link y domain[33], similares en su uso a lo descripto anteriormente para Google. Ixquick permite realizar además, búsquedas de imágenes, video y números telefónicos de personas y empresas, operaciones que hacen que a Ixquick se lo considere también un motor de búsqueda, puesto que en este caso particular, se alimenta de sus propias bases de datos.

Con respecto a The Info (http://www.theinfo.com) cabe mencionar que solo puede utilizar el operador "", y permite la búsqueda de películas, música, noticias, computadoras y libros, entre otros, en diversos motores de búsqueda que trabajan puntualmente con esas temáticas.

Otras posibilidades que ofrecen

Entre otras posibilidades que nos permite realizar este tipo de buscador se tienen:

- *Configurar opciones de funcionamiento:* que permiten la selección del idioma con en el que se trabajará para efectuar las búsquedas, la personalización por parte del usuario de su interfaz (colores de las páginas, cantidad de resultados, etc.) y el seguimiento de sus búsquedas más recientes.
- *Realizar búsquedas en sus propias bases de datos*: que amplían el campo de búsqueda de sus resultados.

---

[32] *Host:* permite efectuar búsquedas en un sitio web determinado, por ej. "radio digital" host:www.clarin.com busca en el sitio de Clarin.com páginas web que contengan la frase "radio digital".

[33] *Domain:* permite efectuar búsquedas en sitios web de un tipo determinado, ya sea educativas (.edu), militares (.mil), gubernamentales (.gov), etc. La búsqueda por ej. de la siguiente palabra clave "radio digital" :.com nos devolverá páginas web comerciales que traten esta temática.

8- La WWW Invisible: definición y características

La WWW "Invisible" o "Profunda" puede definirse como: *"…la porción de la Web que no puede ser indexada*[34] *por los buscadores comunes*".

Compuesta por información, en muchos casos rica y aprovechable en cuanto a los recursos académicos, científicos y generales que puede ofrecer, y en otros casos, peligrosa[35], su acceso no puede ser realizado por los buscadores comunes, ya sea por limitaciones en su funcionamiento interno, como por las restricciones impuestas desde los propios sitios web, que no desean, no pueden o no deben ser visualizados.

Se encuentra clasificada en cuatro componentes de acuerdo a la información que no puede accederse en cada uno de ellos:[36]

- *Web Opaca:* se compone de contenidos que no pueden ser indexados por un buscador por distintas razones:

  -cuestiones de economía: cuando los buscadores almacenan datos de un sitio web en sus bases de datos solo registran algunas páginas de los mismos, no todas. Por otra parte, y a la hora de devolver resultados, solo muestran un fragmento de los mismos, no todos.

  -cuestiones de actualización en los rastreos: los buscadores solo muestran las páginas de los sitios que fueron encontrados en el último rastreo de los robots. Si se agregan nuevos sitios antes del próximo rastreo, estos no serán devueltos hasta una nueva realización del mismo.

  -cuestiones de prestigio o Page Rank: los buscadores solo devuelven páginas de sitios con mayor cantidad de enlaces, referencias o votos. Además, las páginas que no han sido nunca referenciadas o enlazadas no son devueltas como parte de los resultados.

- *Web Privada:* contiene información a la que no puede accederse dado su carácter secreto o confidencial. Como ejemplos se pueden mencionar:

  -páginas protegidas con contraseñas: que permiten acceder al correo electrónico, a la moderación de comentarios, mensajes emitidos en foros de discusión, etc.

---

[34] *Indexar:* Acción de registrar ordenadamente datos e informaciones para elaborar su índice (diccionario de la Real Academia Española: http://www.rae.es/rae.html).

[35] Se dice que existen diferentes niveles de navegación a la hora de recorrer la Web Invisible o Profunda. Los mismos permiten navegar desde sitios limpios y puros hasta los sitios más oscuros y peligrosos.

[36] En la presentación con diapositivas correspondiente a la Clase 3 de esta asignatura puede consultar ejemplos de los distintos casos en los que los buscadores comunes no pueden indexar la información.

-páginas que no desean aparecer como parte de los resultados de un buscador: por lo que se acompañan de un archivo denominado "robots.txt[37]".

-partes del contenido de una página que no conviene que aparezcan en los resultados de un buscador: y que utilizan el comando "noindex[38]".

- *Web Propietaria:* incluye páginas a las que es necesario registrarse para acceder al contenido de forma gratuita. Como ejemplos se puede mencionar el sitio web mexicano de la Secretaría de Educación del Estado de Veracruz (http://rechum.sev.gob.mx:8093/login.aspx) que requiere del registro del usuario para que éste pueda consultar acerca de los servicios que provee y los formatos que puede descargar; el sitio web español de Distribución de Material Veterinario, especializado en Traumatología, Ortopedia y Endoscopia Veterinaria (http://www.vetdmv.com) que permite el acceso a diverso material formativo y productos que se pueden comprar en línea; el sitio de la Asociación de Hoteleros y Gastronómicos de Gualeguaychú, Uruguay (http://www.gyhgualeguaychu.com/) que permite, luego del registro, el acceso a cursos, capacitación, foros y boletín digital; etc.

- *Web Realmente Invisible:* se compone de:

  -archivos de la Web en diversos formatos, ya sea de texto (.txt, .rtf, .doc, etc.) portables (.pdf), de presentaciones con diapositivas (.pps, .ppt, etc.), de animaciones (.swf, .gif, .mng, etc.), comprimidos (.rar, .zip, etc.), dinámicos (.jar, .js, .asp, .php, etc.).

  -páginas que se generan dinámicamente en tiempo de ejecución: como respuesta a una o más acciones del usuario, por ej., ante una consulta a una base de datos, un comentario en una sección de noticias, una participación en un foro de discusión, etc.

  -páginas que envían información en tiempo real: mediante la transmisión en vivo de un programa de televisión, un partido de futbol, el estado del tránsito de automóviles o subterráneos, etc.

---

[37] Existe mucha información en la Web que nos indica cómo configurar este archivo, de las que se pueden mencionar a: http://www.neoteo.com/crear-y-personalizar-el-archivo-robot-txt/ y http://www.emezeta.com/articulos/robots-txt-todo-lo-que-deberia-saber. También existen páginas de buscadores que nos indican cómo trabajan los robots en relación al mismo: https://support.google.com/blogger/answer/2472627?hl=es-419

[38] En el siguiente enlace: http://www.entrecodigos.com/2007/11/robots-nofollow-noindex.html se explica cómo trabaja este comando y se realiza una comparativa con el archivo: "robots.txt".

Entre las principales herramientas informáticas que se pueden utilizar para encontrar información en la WWW Invisible o Profunda de carácter rico y aprovechable, se tienen los buscadores especializados por directorio y los motores de búsqueda. Estos buscadores, al igual que en el caso de la WWW Visible o Superficial, nos permiten la realización de operaciones básicas y avanzadas con las que es posible acceder a la diversidad de los tipos de archivos que la componen.

9- Buscadores por directorio, temas o categorías en la Web Invisible o Profunda

Construyen sus bases de datos en base a los aportes de personas o robots que toman como fuente de información contenidos de otras bases de datos de carácter científico, médico, académico, económico, militar, educativo, etc.

Como ejemplos de este tipo de buscadores se pueden mencionar:

- *a nivel mundial y generalista:* Science Direct (http://www.sciencedirect.com/) una base de datos científica de artículos de revistas que ofrece textos completos y capítulos de libros de casi 2.500 revistas y 26.000 libros.
- *a nivel mundial y especializado*: Comunicación Invisible.net (http://www.comunicacioninvisible.net/wai/) que provee la descripción y el acceso a bases de datos de dominio público en la Web, relacionadas con estudios fílmicos. Creada por españoles, se dirige a investigadores, docentes y profesionales en el área de la comunicación, cinematografía y la documentación audiovisual.
- *a nivel regional y especializado:* Infomine (http://infomine.ucr.edu/) que provee acceso a una gran diversidad de librerías de la Universidad de California, EE UU, con información relativa a medicina, economía, cultura y gobierno, entre otros.

Características y posibilidades que ofrecen

Al tratarse de buscadores por directorio, la lógica de búsqueda depende de si el usuario que la ejecuta conoce o no, el tema o la categoría en la que podría encontrarse la información que está buscando, de modo que la forma de buscar es similar a la que se produce en la Web Visible, con la diferencia de que aquí se manipulan bases de datos.

¿Cómo participar?

Algunos buscadores temáticos de la Web Invisible permiten que el usuario participe en la elaboración de sus bases de datos mediante la realización de sugerencias o ediciones de nuevos sitios web, mediante las categorías de usuarios "común" y "editor", descriptas en la teoría de la Clase 1.

Otras posibilidades que ofrecen

Los buscadores por directorio de la Web Invisible permiten la realización de diversas operaciones relacionadas con la información que devuelven. Entre ellas se puede mencionar:

- *Búsquedas completas y detalladas de los recursos de las bases de datos:* mediante consultas que permiten especificar categoría o tema, autor, clasificación, descripción, editor, palabras clave, texto, título y URL, entre otros.
- *Especificar las búsquedas por palabra clave mediante la provisión de operadores de búsqueda:* que permiten establecer uno o más filtros de selección de resultados, como se puede apreciar a través de la búsqueda avanzada de algunos buscadores[39].
- *Configurar el tipo de recursos en los que se buscará la información:* ya sea en artículos de bases de datos, bibliografías, catálogos de libros, páginas gubernamentales, imágenes, video, software, etc.[40]
- *Configurar el modo de presentación y organización de los resultados:* a través de la selección de la cantidad de resultados por página, el lugar de su visualización y el modo de visibilidad, ya sea por fecha, por orden alfabético, por una letra en particular o por orden de importancia, entre otros.
- *Realizar búsquedas en la Web Visible:* a través del acceso a listas de emails, directorios telefónicos, grupos de noticias, páginas amarillas y motores de búsqueda, entre otros.

10- Motores de búsqueda o Buscadores por palabra clave en la Web Invisible o Profunda

[39] *Comunicación Invisible.net* (http://www.comunicacioninvisible.net/wai/index.php?P=Advanced) permite la creación de filtros compuestos de búsqueda mediante el agregado de uno o más filtros simples.

[40] *Infomine* permite que pueda especificarse el lugar de la búsqueda a partir de sus principales recursos http://infomine.ucr.edu/cgi-bin/browse?category=maps&field=resource_types

Constituyen una fuente de información y acceso a base de datos poderosa y potente, en el que las búsquedas se producen fundamentalmente a través de palabras clave, construidas en base al uso de uno o más operadores.

Como ejemplos de este tipo de buscadores se pueden mencionar:

- *a nivel mundial y generalista:* Wayback Machine (http://archive.org/web/web.php) que permite la navegación a través de más de 240.000.000 de páginas web con información general y de interés, escritas en diversos idiomas y archivadas desde el año 1996 en adelante.
- *a nivel mundial y especializado*: Science Direct (http://www.sciencedirect.com/) una de las herramientas de investigación científica más completa de la Web tal como se mencionó anteriormente.
- *a nivel regional y especializado:* Plants (http://plants.usda.gov/java/) provee información estandarizada acerca de plantas vasculares, musgos, hepáticas, antocerotas y líquenes de los EE.UU. y sus territorios.

¿Cómo buscar?

Al igual que con los motores de búsqueda de la Web Visible el usuario coloca la palabra clave de búsqueda, haciendo uso de uno o más operadores.

¿Cómo especificar la búsqueda?

Los motores en la Web Invisible permiten la realización de búsquedas completas a través del uso de operadores y de consultas que pueden realizarse sobre los diferentes campos de las bases de datos.

En relación a los operadores, la mayor parte de ellos ya fueron descriptos cuando se abordaron los motores de la Web Visible. Con respecto a las consultas, estas se constituyen en una modalidad de búsqueda muy potente e interesante por las diferentes posibilidades que pueden llegar a abarcar.

Entre los principales operadores utilizados se tienen:

- *Operador "":* se utiliza cuando se desea buscar una palabra o frase exacta en las bases de datos.
- *Operador OR:* busca una o más palabras claves a la vez.
- *Operador AND:* busca todas las palabras claves a la vez.
- *Operador -:* excluye una o más palabras clave de los resultados de una búsqueda.
- *Operador *:* busca cero o más palabras claves relacionadas a una búsqueda.

- *Operador ANDNOT:* constituye una combinación de los operadores AND y – mencionados anteriormente.

Con respecto a las consultas estas pueden efectuarse de acuerdo a los siguientes parámetros:

-*Autor del recurso:* institución, organismo, asociación, empresa, etc.

-Título: nombre o título del recurso a buscar.

-*Descripción del recurso:* breve resumen que indica de lo que trata el mismo.

-Palabras clave: que permitan etiquetar el/los temas de los que trata el recurso.

-*Fecha:* antes, después o entre un período de años determinado.

-*Tipo de archivo:* .pdf, .html, .doc, cualquier formato, etc.

-*Origen del recurso:* si fue indexado por una persona o robot.

-*Acceso del recurso:* gratuito o basado en una suscripción.

-*Tipo de recurso a buscar:* resúmenes, artículos, artículos en periódicos, libros, conferencias, patentes, opiniones, páginas científicas, tesis y disertaciones

-*Lugar en el que se busca la información:* bases de datos, universidades, periódicos en línea, revistas especializadas, etc.

-*Tipos de resultados a mostrar:* de agricultura, astronomía, química, computación, de negocios, información ambiental, ingeniería, economía, etc.

Otras posibilidades que ofrecen

Además de las posibilidades que se obtienen del uso de los operadores y parámetros anteriormente mencionados, los motores permiten:

- *Realizar búsquedas por diversas categorías de información:* como si se tratase de un buscador por directorio[41].
- *Efectuar búsquedas completas de tipologías concretas de información*: tales como video, textos, audios, proyectos, noticias de televisión, etc.[42].
- *Configurar su funcionamiento:* en cuanto a la selección de la cantidad de resultados por página que pueden visualizar y el lugar y el orden de

[41] El motor de búsqueda Plants (http://plants.usda.gov/java/) brinda una posibilidad conjunta de buscar información tanto por palabra clave como así también por categorías.
[42] En el sitio de Wayback Machine (http://archive.org/about/) se puede realizar este tipo de búsquedas.

preferencia (fecha, relevancia, por autor, etc.) en el que estos serán devueltos, entre otros.

11- Buscadores de bases de datos en la Web Invisible o Profunda

Si bien el mayor predominio de los tipos de buscadores existentes en la Web Invisible son motores o directorios, es posible encontrar buscadores de bases de datos en los que podría encontrarse la información que buscamos.

¿Cómo buscar en un buscador de bases de datos?

La lógica de búsqueda puede diferir entre un sitio y otro, por lo que es recomendable explorar a fondo c/u de ellos para apropiarse de su manejo. De todas formas, es posible establecer para el usuario la siguiente secuencia de estrategias:

1) Selecciona el lugar donde colocará la “palabra clave”, pudiendo elegir buscar en uno o en más sitios web de bases de datos proporcionadas por el buscador.

2) Ejecuta la búsqueda a través del botón de Búsqueda.

3) Examina los resultados donde puede encontrarse directamente lo que está buscando, o bien, los accesos a las bases de datos.

Como ejemplo de este tipo de buscador se puede mencionar a LANIC: Latin American Network Information Center de la University of Texas (http://lanic.utexas.edu) que permite acceder a archivos y bases de datos académicas relacionadas con Latinoamérica.

12- Otros buscadores de la Web Invisible

Además de la tipología de buscadores mencionada anteriormente existen numerosos sitios de gobierno, universidades, de organizaciones no gubernamentales, etc. que permiten el acceso a diversidad de información de carácter público contenida en sus páginas o bases de datos. Como ejemplos se puede mencionar el Sistema Argentino de Información Jurídica (http://www.infojus.gov.ar/) que permite realizar búsquedas de leyes, decretos nacionales, sentencias, dictámenes y normas de nuestro país o la Galería Nacional de Arte de los EE UU

(http://www.nga.gov/content/ngaweb/Collection.html) que permite navegar por sus pinturas, esculturas, dibujos y fotografías digitalizadas.

Diversos autores, tales como Crucianelli (2010) han realizado una recopilación muy importante de estos recursos que se pueden visitar y explorar en profundidad. También se pueden realizar búsquedas en los motores, directorios o meta/multi buscadores de la Web Visible para encontrarlos.

13- La WWW Semántica: definición y características

La WWW Semántica surge como parte de la evolución actual de Internet denominada Web 3.0 y se define como: "*una extensión de la Web existente que permite la realización de búsquedas más inteligentes a través de buscadores inteligentes*".

Concebida originalmente por Tim Berners Lee con la idea de mejorar las búsquedas de información, de forma tal que las máquinas no solo fueran capaces de utilizar información sino también interpretarla, constituye apenas una pequeña porción del 100% de la WWW, de la que se estima que ocupa un 8% contra un 17% de la Web Superficial y un 75% de la Web Invisible y fue

Su propuesta es la de ayudar a los usuarios, encontrando la información por ellos de forma rápida y con un mayor porcentaje de acierto. Para ello necesita asociar un contenido semántico a la información que circula por la Web, compuesto por los siguientes elementos:

- *RDF:* permite describir los recursos de la WWW (páginas web y archivos en diversos formatos) a través de los metadatos[43] especificados en lenguaje XML[44].
- *SPARQL:* posibilita la realización de búsquedas inteligentes sobre los recursos de la Web Semántica utilizando distintas fuentes de datos. Estas búsquedas son efectuadas por programas especiales denominados "agentes inteligentes" y tienen la capacidad de comprender las consultas realizadas por los usuarios.
- *OWL:* se utiliza para definir ontologías o relaciones entre los conceptos que se asocian a cada página de la WWW. Las ontologías son utilizadas por los usuarios, las bases de datos y las aplicaciones que necesitan compartir información específica.

---

[43] *Metadatos:* etiquetas que se asocian a un contenido y que permiten describir diferentes características del mismo, a saber su autor, su título, su descripción, el contexto en el que puede ser utilizado, etc.

[44] *XML:* Extensible Markup Language (lenguaje de marcas extensible) que permite etiquetar un contenido.

14- Buscadores de La WWW Semántica: definición y características

Con el agregado de contenido semántico a la información se facilita que la misma tenga una cierta dosis de inteligencia artificial que les permite a los buscadores especializados en este tipo de web "comprender" qué es lo que realmente busca el usuario.

Existen algunos proyectos de buscadores semánticos, que se encuentran funcionando en versiones beta[45], cada uno con su propia lógica y funcionamiento de búsqueda, si bien pueden encontrarse algunas similitudes en cuanto a los operadores de búsqueda que proveen.

Al respecto es importante mencionar los siguientes buscadores:

- *Wolfram Alpha* (http://www.wolframalpha.com/) disponible en inglés y el que se pueden realizar búsquedas de información a través de conceptos[46] o diversos operadores.
- *Swoogle* (http://swoogle.umbc.edu) que posibilita la realización de búsquedas por ontología (relaciones entre conceptos) o documentos.
- *Twine* **(**http://www.twine.com/) que permite la localización de personas con intereses comunes así también como la localización de temas específicos.

Operadores de búsqueda

Existen diversos operadores de búsqueda y estos varían de un buscador a otro. En el caso particular del buscador Wolfram Alpha se pueden utilizar los siguientes:

-*"vs" (versus):* permite efectuar una comparación entre dos conceptos o términos que pertenecen a la misma ontología, por ej., de compañías o frutas, etc. Cuando se comparan compañías se pueden apreciar por ej. sus estadísticas financieras de manera anual, mientras que si se comparan frutas, pueden observarse sus diferentes informaciones nutricionales.

-*"Where am i" (donde estoy):* permite rastrear la computadora en la que nos encontramos ubicados a partir de la localización de su dirección IP y las coordenadas geográficas de su ubicación física.

---

[45] *Versión Beta:* versión de software que ha pasado la etapa de prueba interna, llamada alfa y que se ha lanzado a los usuarios para la realización de pruebas públicas.

[46] Por ejemplo, si se ingresa la palabra Brasil, Wolfram Alpha realizará un informe sobre el nombre completo del país en su forma nativa, su bandera, su ubicación geográfica en el mundo, sus datos económicos, demográficos e incluso una tabla de conversión de valores monetarios.

-*“Am I drunk” (¿estoy borracho?):* permite llenar un formulario para saber en el caso de haber bebido, las condiciones de salud en las que uno se encuentra.

-*“body mass index” (índice de masa corporal):* permite ver en base a los datos de altura y peso que se ingresen en un formulario si estos están dentro del rango esperado.

-*“8 character password” (generador de contraseñas):* permite generar contraseñas con combinaciones de letras y números.

El resto de los buscadores posee sus propios operadores, por lo que se deberá familiarizar con los mismos a través de los menús de ayuda o los resúmenes explicativos que proveen.

15 Comparando la Web Superficial con la Web Profunda y Semántica

Las Webs actuales (Superficial y Profunda) poseen una gran capacidad para almacenar datos, leer y visualizar sus contenidos, pero no son capaces de entender todo lo que contienen, algo que la Web Semántica sí es capaz de hacer. La automatización de los procesos en los casos que se desee hacer por ej., la reserva de un vuelo o la consulta de los medios de transporte disponibles en una ciudad para trasladarse por la misma es favorecida por los servicios provistos por los buscadores semánticos y los buscadores o agentes computarizados inteligentes que no solo rastrean e indexan la información, sino que también tienen la capacidad de interpretarla, pudiendo interactuar con el entorno para el cumplimiento de las tareas encomendadas.

La búsqueda de una información puntual, por ej. “viajes a Perú, Lima” permite devolver resultados más inteligentes, por ej. en un buscador como Wolfram Alpha, que devolverá, entre otros datos, el tiempo que se tarda en llegar mediante un vehículo en particular y la cantidad de kilómetros de distancia a los que nos encontramos. Sin embargo, la Web Semántica, aún se encuentra en desarrollo y contrario a lo que muchos profetizaron años atrás, aún faltan algunos años para alcanzarla en su totalidad.

Existen muchos proyectos que han puesto en marcha el desarrollo de aplicaciones para favorecer a la Web Semántica, accesibles desde el World Wide Web Consortium (http://www.w3.org/) una comunidad internacional donde los miembros de las organizaciones, su staff completo y el público trabajan para desarrollar su estándar web. Pero aún definiendo los valores RDF, SPARQL y OWL se debe continuar trabajando en los metaprogramas, así también como en los servicios semánticos.

16- Bibliografía de interés

- Crucianelli, S. (2010). Capítulo 1: “Searching”, del libro: *Herramientas digitales para periodistas*. Iniciativa del Centro Knight para el Periodismo en las Américas de la Universidad de Texas. Disponible en: https://knightcenter.utexas.edu/hdpp.pdf
- Crucianelli, S. (2010). Capítulo 3: “Documentos oficiales y acceso a bases de datos” y Capítulo 7: “La Web Semántica”, del libro: *Herramientas digitales para periodistas*. Iniciativa del Centro Knight para el Periodismo en las Américas de la Universidad de Texas. Disponible en: https://knightcenter.utexas.edu/hdpp.pdf
- González, M. J & otros (2003). Periodismo de Investigación: los medios recuperan su rol. Disponible en: http://www.archivochile.com/tesis/10_tmdec/10tmdec0001.pdf
- Jaramillo, M. (2011). Google para periodistas. Disponible en: http://www.slideshare.net/EDWINLOZADA/google-para-periodistas
- Stallman, R. & otros (2003). Contra el Copyright. Disponible en: http://www.tumbonaediciones.com/vs-copyright.pdf
- World Wide Web Consortium (2013): “Guía breve de la Web Semántica”. Disponible en: http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica