

INSPT - UTN

COMPUTACION

BUSCADORES

(Internet)

Los **objetivos** de esta clase son conocer las herramientas que Internet ofrece para diseñar estrategias de navegación e interrogación; observar y reconocer las características de estas herramientas sin perder de vista sus diferencias, para poder determinar en qué casos utilizar unas u otras. Además, intentaremos abordar el concepto de Internet invisible, para conocer su alcance y características.

Herramientas de búsqueda en Internet

Cuando sentimos la necesidad de obtener alguna información, lo que generalmente solemos hacer como primer impulso es recurrir a un buscador de Internet para satisfacerla, en general lo hacemos con Google. Si bien en muchas ocasiones este recurso puede resultar suficiente, es preciso tener en cuenta que hay mucha información importante que no aflora en las búsquedas convencionales y que mucha de la que aparece únicamente aporta ruido y confusión (Cordón García, et. al. 2012)

En Internet existen diferentes tipos de herramientas de búsqueda, en esta clase nos centraremos en estudiar las más habituales:

- **Directorios**
- **Metabuscadore**s
- **Motores de búsqueda académicos**
- **Buscadores semánticos**

Directorios

Los directorios son aplicaciones controladas por humanos que manejan grandes bases de datos con direcciones de páginas Web. Estas bases de datos son alimentadas cuando los administradores revisan las direcciones recibidas que luego clasifican en subdirectorios. A su vez, están organizados en categorías temáticas, estructurados jerárquicamente en un árbol de materias de información que permite visualizar los recursos descendiendo desde los temas más generales a los más específicos. Cada categoría presenta un listado de enlaces que en algunos directorios ofrecen una descripción breve de su contenido.

Es decir que los directorios son guías o listas agrupadas y ordenadas sistemáticamente por categorías y subcategorías que registran las direcciones y una pequeña descripción de los diferentes sitios o recursos disponibles en Internet.

Podemos navegar los directorios de dos maneras:

- Explorando la estructura de las categorías temáticas.
- Buscando por palabra clave sobre el conjunto de referencias contenidas.

¿Qué información podemos buscar en un directorio?

El uso de directorios para una búsqueda de información reviste indudables ventajas, como por ejemplo su facilidad de uso, la selección de sitios que realizan, lo que en muchos casos implica un valor añadido de calidad y la agrupación por categorías, que facilita la localización de temas de interés. Pero tienen el inconveniente de que funcionan sobre bases de datos reducidas y, sobre todo, de que sus actualizaciones no son lo rápidas que sería deseable (Cordón García, 2012).

Una de las iniciativas más interesantes en el ámbito de los directorios en Internet es la representada por **Dmoz**: un proyecto colaborativo en el que editores voluntarios listan y categorizan enlaces a páginas Web. Cualquier

persona puede sugerir un enlace en una categoría determinada que luego ha de ser aprobada por un editor, también cualquier interesado puede ofrecerse como editor completando un formulario en el que defiende su aptitud para organizar una categoría.



<http://www.dmoz.org/>

Motores de búsqueda

¿Qué características distinguen a un directorio de un motor de búsqueda?

Los motores de búsqueda, ya sea **Google**, **Yahoo** o **Bing**, no cuentan con subcategorías como los directorios, sino con avanzados algoritmos de búsqueda. Los motores son aplicaciones que manejan grandes bases de datos de referencias a páginas web recopiladas por medio de un proceso automático, sin intervención humana.

Pero ¿en qué varía la información que provee un buscador en relación a la que provee un directorio? ¿Bajo qué circunstancia se utiliza uno u otro?

Los directorios son muy útiles si no sabemos cómo “describir” lo que buscamos, o buscamos palabras con varios significados o necesitamos una mejor comprensión del tema.


Por ejemplo, si buscamos en un directorio "fotografía", nos encontraremos con un listado de sitios ordenados y seleccionados que van desde "historia de la fotografía" hasta "tipos de cámaras fotográficas", junto a temas como "cine" o "artes plásticas"

Los resultados de búsqueda de "fotografías" obtenidos con un motor de búsqueda ofrecerán una visión mucho más desordenada y menos precisa del tema.

Metabuscadores

Otra herramienta factible para buscar información en Internet son los **metabuscadores**, que tienen una interfaz de búsqueda que consulta simultáneamente varios motores y permite navegar al mismo tiempo varias bases de datos. Estos sistemas no almacenan direcciones y descripciones de páginas en su base de datos (diferencia con los Directorios), pero contienen registros de motores de búsqueda e información sobre ellos.

Por ejemplo: cuando usamos un metabuscador, el sistema envía nuestra petición a todos los motores de búsqueda que tiene registrados (Google, Yahoo, etc.). La principal ventaja reside en la capacidad de combinar los resultados de muchas fuentes y en el hecho de que podemos acceder a varias fuentes de información en forma simultánea a través de una interfaz sencilla y amigable. Entre los más interesantes se encuentran los dos que se detallan a continuación.

	<p>Ixquick: la búsqueda se realiza simultáneamente en varios motores de forma combinada dando relevancia a los primeros diez resultados. Usa un sistema de estrellas para calificar dichos resultados, es decir coloca una estrella por cada resultado devuelto dentro de los diez primeros y los identifica por motor de búsqueda. Por tanto, los resultados destacados son aquellos que han sido devueltos por la mayoría de los motores de búsqueda. También permite buscar en 17 idiomas diferentes. https://ixquick.com/esp/</p>
---	---

	<p>Dogpile: la búsqueda se realiza simultáneamente en varios motores. En los resultados aparecen sugerencias para proseguir la indagación por áreas o temas conexos. La mascota de Dogpile es un perro dálmata llamado Arfie, el cual aparece en ocasiones haciendo ciertos festejos de acuerdo a los resultados de búsqueda obtenidos. http://www.dogpile.com/</p>
--	--

Motores de búsqueda académica

Los motores de búsqueda académica permiten buscar contenidos de calidad, filtrados y fiables, como artículos científicos, tesis doctorales o informes técnicos alcanzando un amplio espectro de temas. También arrojan datos sobre el rendimiento o la proyección de un trabajo en cuestión indicando los trabajos que lo citan y las menciones que recibe. El motor más conocido es Google Académico o Google Scholar: <http://scholar.google.com.ar/>

Google Académico es un buscador de publicaciones científicas y un índice de citas que ayuda a conocer el impacto que las publicaciones tienen. Esta última propiedad es la que más interesa y la que lo convierte en una competencia directa de otros índices de citación como **Web of Science (WoS)** de Thomson Reuters, o **Scopus**, de Elsevier (Torres Salinas, 2009). Además permite obtener las referencias de los artículos e incorporarlas a un gestor bibliográfico. Es una herramienta ideal para realizar al menos tres tareas:

- Buscar el texto completo de un trabajo.
- Buscar la producción bibliográfica de un autor, de una revista o sobre un tema.
- Buscar las citas que recibe un trabajo (libro, artículo de revista, tesis, informe)

Otra de las particularidades fundamentales es su gratuidad, marcando una distancia enorme con el resto de los proveedores, más si tenemos en cuenta el elevado precio de las licencias de las bases de datos, fomentando el acceso universal a la información científica, y favoreciendo el incremento de la presencia de publicaciones científicas en acceso abierto.

Les dejamos 1 video realizado por la **Biblioteca de la Universidad de Castilla La Mancha** dónde muestran en forma sencilla las funcionalidades de Google Académico:

<http://youtu.be/22W5gt28flg>

Buscadores semánticos

Un motor de búsqueda semántica intenta dar sentido a los resultados en función del significado de los términos y del contexto en donde se busca la información. Por ejemplo, si buscamos "**elección**", en un buscador semántico podríamos recuperar documentos que contienen la palabra "**voto**" o "**campaña**", aunque la palabra "**elección**" no se encuentre en los documentos recuperados.

Una parte importante de este proceso es la "desambiguación" del significado de la palabra que permite identificar el sentido con el que la usamos en un determinado contexto. Esto sucede cuando dicha palabra tiene pluralidad de significados, es decir, es un término polisémico.

Los bibliotecarios estamos familiarizados con la desambiguación y la polisemia de los términos; por ejemplo, cuando estudiamos los tesauros, tenemos dos formas de desambiguación: los **calificadores** y las **notas de alcance**.

- El **calificador** es un término que se agrega entre paréntesis al descriptor (y forma parte de este) a fin de distinguir entre distintos homónimos o términos polisémicos dentro del tesauro, por ejemplo, *tanques (contenedores)* y *tanques (vehículos)*.
- Las **notas de alcance** son descripciones que permiten definir el término.

Tomemos un ejemplo de **desambiguación** en Wikipedia: si buscamos el término "**vela**", vemos que aparece una página de desambiguación donde indica que "vela" puede referirse a: **vela de cera o combustible para la iluminación** o **vela de barco** para aprovechar la energía del viento. Esto permite y facilita que el usuario encuentre la información que se adecua a su necesidad.



<http://es.wikipedia.org/wiki/Vela>

En suma, los buscadores semánticos seleccionan por nosotros. Eligen los documentos que necesitamos recuperar porque "entienden" el contexto de nuestra búsqueda y "entienden" lo que significan las palabras. Es decir, que si nosotros estamos interesados en la navegación y buscamos sobre esa temática, el buscador

semántico entenderá que cuando buscamos “vela”, se refiere a la vela de los barcos y evitará traernos los resultados que contengan “vela” para iluminar.

Podemos decir que los buscadores semánticos mejoran la búsqueda web tradicional, pero no la reemplazan. Una gran parte de las consultas se realiza por medio de la navegación, y la búsqueda semántica no es un sustituto para ello. Por otra parte, las búsquedas específicas se beneficiarán con la búsqueda semántica, aunque aún falta mucho por mejorar en los buscadores de ese tipo.

Internet Invisible

¿Alguna vez se preguntaron si toda la información que está en la Web es recuperada cuando hacen una búsqueda?

Internet invisible o Internet profunda (en **inglés**: *Deepnet*, *Invisible Web*, "Deep web", *Dark Web* o *Hidden Web*) es justamente esa información que no es recuperada a través de los buscadores convencionales.

Dentro de Internet invisible, se pueden agrupar los siguientes recursos o herramientas y toda la información que ellos contienen:

- Catálogos de bibliotecas y bases de datos bibliográficas (OPAC y las bases de datos de registros bibliográficos accesibles a través de pasarelas web)
- Bases de datos no bibliográficas (bases alfanuméricas, a texto completo, diccionarios y/o enciclopedias)
- Revistas electrónicas, en las que es necesario un registro previo y las que solo se puede recuperar la información mediante búsquedas en su base de datos
- Documentos en formatos no indizable, como documentos en pdf, word...
- Obras de referencia: enciclopedias, diccionarios... en las que es necesario interrogar a la base de datos para acceder al contenido.

Los buscadores convencionales recorren las páginas Web e indizan su contenido. Sin embargo, los programas que utilizan para hacerlo no pueden extraer la información de algunas bases de datos, ya que ésta es sólo recuperada dinámicamente, en respuesta a una consulta específica. Entonces, para recuperar la información que hay estas bases, es necesario conocerlas y acceder directamente.

Causas

La principal causa de la existencia de la *Internet Invisible* es la imposibilidad de los motores de búsqueda de encontrar o **indexar** el 95% de la información existente en Internet. Si los buscadores pudieran acceder a toda la información entonces la "Internet profunda" desaparecería, pero esto es imposible porque siempre existirán páginas privadas.

Los siguientes son algunos de los motivos por los que los buscadores son incapaces de indexar la *Internet Invisible*:

- Páginas protegidas con contraseña
- Documentos en formatos no indexables
- Enciclopedias, diccionarios, revistas en las que para acceder a la información hay que interrogar a la base de datos

Un poco de historia del término...

Jill Ellsworth en 1994, doctora especializada en el estudio de internet, empleó el término “internet invisible” para referirse a la información que no podían encontrar los buscadores más comunes (Ellsworth, 1995). En el año 2000, un estudio elaborado por Michael Bergman (Bergman, 2001), ratificaba la presencia de una “red profunda” de aproximadamente unos 7500 terabytes (7500 billones de bytes) frente a los 19 de la web visible accesible por medio de los buscadores convencionales.

Codina opina que Internet invisible “es un nombre claramente inadecuado para referirse al sector de sitios y de páginas web que no pueden indizar los motores de búsqueda de uso público como Google o AltaVista. Propone web “no indizable”, que en su opinión “un término mucho más adecuado, pero claramente alejado de la capacidad de sugeridora del término invisible” (Codina,2003).

<http://www.cronicaweb.com/deep-web-lo-conoces-3753/>