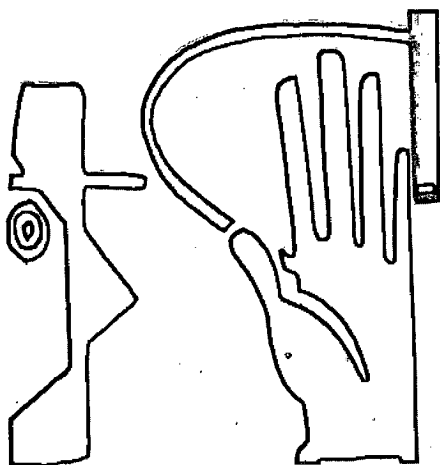


Internet como nueva herramienta de investigación

Encuentre agujas en un pajar

□ Zinnia Martínez Araujo

Para el investigador de la comunicación y de todas las áreas del saber, Internet es la herramienta de información más poderosa hasta ahora inventada; sin embargo, para muchos, la sensación de ansiedad y frustración es muy grande cuando sentimos que nuestras búsquedas no nos llevan a nada y que el volumen de información y de basura nos avasalla. Tranquilícese, hablele a su computadora en su idioma y verá como a veces es posible encontrar una aguja en un pajar.



Internet se ha convertido para el Comunicador Social de esta era, no sólo en un nuevo recurso que rompe con una cantidad de barreras impensables en los medios tradicionales, la interactividad, la universalidad, la conjunción de varios medios; sino en un reto profesional que amplía nuestro trabajo, nuestros lectores, nuestras fuentes hasta límites insospechados.

La relación informante-informado ha cambiado para siempre, el flujo de la comunicación ha empezado a saltarse algunos eslabones de la cadena y la investigación puede actualmente llevarnos allí donde surge el documento inicial, la fuente primera, la versión local multiplicada por cientos y cientos de versiones en esta gran red donde todo el mundo tiene algo que decir.

Son por todos conocidas las maravillosas ventajas que supone Internet para la distribución y hallazgo del conocimiento. Sabemos lo fabuloso que es poder acceder a la información global desde nuestras casas y encontrarnos con todo un mundo de documentos e informaciones valiosísimas que hace unos años no hubiéramos ni pensado en poder tener al alcance de nuestras manos.

Pero también sabemos, que la red de redes crece a pasos acelerados e incontrolables cada día y que encontrar realmente lo que se busca no es una tarea fácil, de hecho, es mucho más complicado de lo que uno sospecha.

Se puede decir que de un tiempo para acá, resulta básico tener una cultura «informativa» que nos permita tener los conocimientos básicos en el uso de una computadora y sus recursos asociados.

De igual forma todo profesional preocupado por la investigación y/o la explotación de Internet como nueva fuente debe aprender a convertir sus necesidades de información en ecuaciones o fórmulas de interrogación que le ayuden a hallar su agujita en este gran pajar que es la red de redes.

Roberto Hernández Montoya¹ afirma que «el periodista-bibliotecario aprenderá como nadie a buscar y sobre todo a encontrar y destilar la información para que sea más útil a quien la necesita.... Su nuevo trabajo será el de guía, cicerone, Baedeker, organizador, administrador de recursos».

Pero para ello debemos empezar por conocer con qué contamos para enfrentarnos ante esto y sacar lo mejor de nuestro trabajo y tiempo invertido. En concreto un profesional de la comunicación debe aprender a formular ecuaciones de búsqueda, así como, conocer las diferencias entre directorios, buscadores, metadirectorios y metabuscadores.

Nuestra intención es proporcionar algunas nociones de lo que son estas herramientas y para qué utilizarlas en concreto, sin ánimos de convertir al lector de este artículo en un gran experto en la búsqueda a través de Internet pero intentando despertar la curiosidad de los interesados y quién sabe si lograr grandes hallazgos.

CONVIERTA SUS PREGUNTAS EN ECUACIONES

Una ecuación de búsqueda es una representación de una necesidad de información. Más de una vez nos habremos visto en la disyuntiva de no saber expresar en los términos adecuados lo que quere-

ALGUNOS WEBS DE INTERÉS PARA PERIODISTAS

SALA DE PRENSA.COM. Es una página dedicada a los profesionales de la comunicación con cientos de recursos de interés, pasando por los web oficiales de algunos organismos gubernamentales del mundo como los links a todos los periódicos de Latinoamérica y el mundo. www.saladeprensa.com

PERIODISMO EN AMÉRICA LATINA. Colección de enlaces con fuentes de información en inglés, español y portugués en todo Latinoamérica. www.lanic.utexas.edu/la/region/journalism

AMERICAN JOURNALISM REVIEW. Seguramente el mayor y mejor centro de recursos sobre el periodismo. Además de los artículos escritos por especialistas, contiene una rica selección de enlaces. www.ajr.org • www.newslink.org

COLUMBIA JOURNALISM REVIEW. Una de las mejores escuelas de periodismo del mundo publica una revista electrónica. En ella colaboran brillantes columnistas. www.cjr.org

PERIODISTAS DE INVESTIGACIÓN. Grupo de periodistas y comunicólogos mexicanos, con vínculos en otros países de Latinoamérica. www.investigacion.org.mx/boletin.html

REPORTEROS SIN FRONTERAS. La ONG de periodistas franceses que actualmente se dedica a velar por el respeto a la libertad de expresión en el mundo entero. www.rsf.fr

PULSO. La revista *Pulso* lleva un reporte de los temas más actuales en Latinoamérica y suma un gran cantidad de enlaces y fuentes muy interesantes. www.pulso.org

PREMIOS PULITZER. Un repaso a los célebres premios Pulitzer y a la trayectoria de sus ganadores. www.pulitzer.org/navigation

PRENSA LATINA. Un Kiosko que recoge los enlaces a todos los medios de comunicación en línea divididos por países. Incluye América Latina y España. www.mundolatino.com

KIOSKO INTERNET. Un índice de todos los medios de comunicación españoles y algunos internacionales. www.kiosko.com

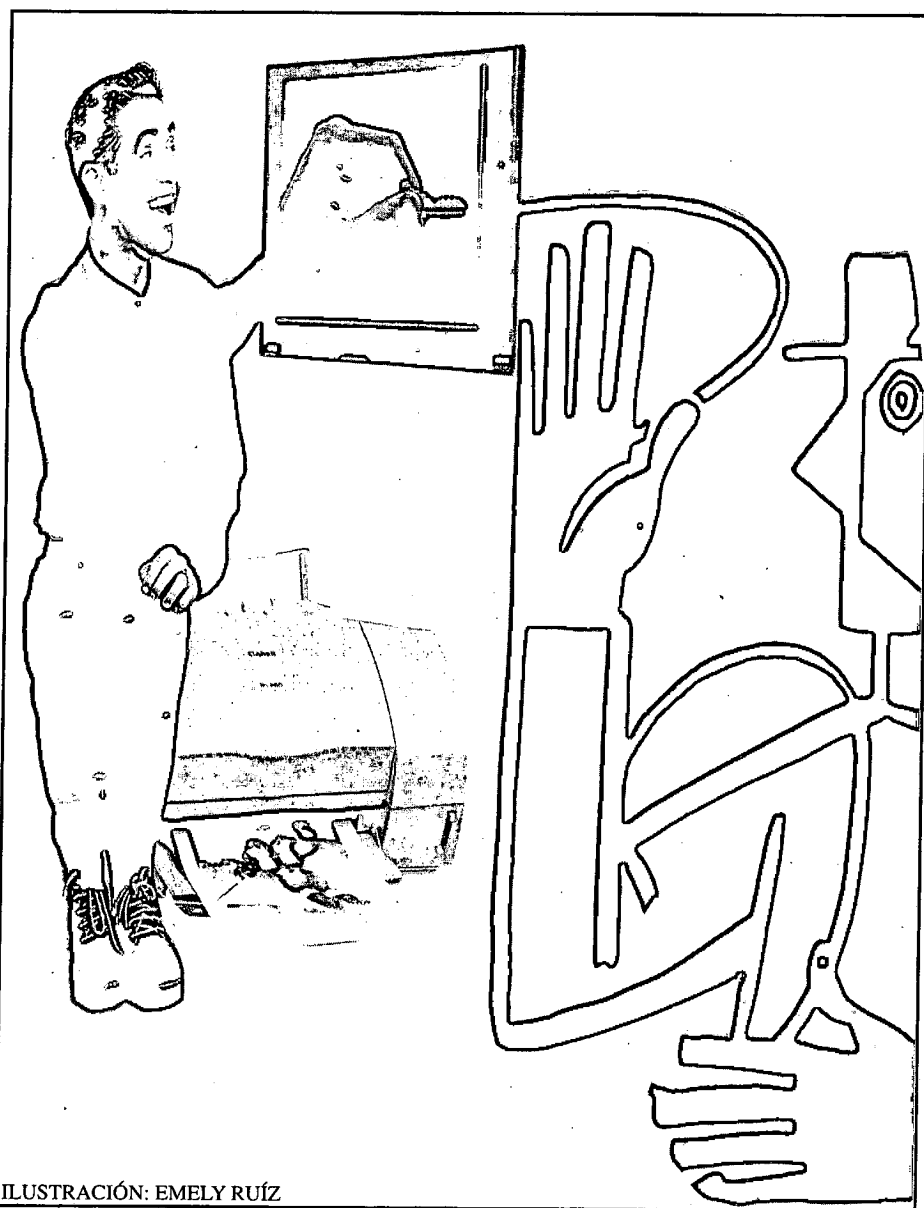


ILUSTRACIÓN: EMELY RUÍZ

mos buscar y eso se nos refleja en millares de documentos que no tienen nada que ver con lo que estamos buscando.

Para evitar eso, existen dos grandes tipos de ecuaciones de búsqueda: las que sirven para representar búsquedas simples y las que sirven para representar búsquedas compuestas, también denominadas ecuaciones booleanas.

Las búsquedas simples implican el uso de un solo término de búsqueda, que puede consistir en una palabra, un nombre, una frase <«juegos olímpicos»> o en la negación de un término <not «juegos olímpicos»> y en cada caso se presentarán los documentos que satisfacen la búsqueda y los que no.

Por su parte, las ecuaciones compuestas requieren de lo que llamamos los operadores booleanos. Los más comunes son las palabras en inglés <and>, <or>, <not> y cada uno de ellos supone la inclusión o exclusión de documentos en nuestra búsqueda.

El operador <and> se utiliza para encadenar dos términos que queremos aparezcan en nuestra búsqueda y se podría traducir como <todos los documentos que incluyan estas dos palabras>

Por otro lado <or> quien también cumple la función de encadenar, se podría traducir como <alguno de estos dos términos>.

Por último el operador <not> sirve para eliminar aquellos términos que no queremos estén en los documentos encontrados. Es decir, utilizar la palabra <not> puede servirnos para aclarar a qué refiere el tipo de documento que queremos.

Por ejemplo: (hipertexto AND diseño NOT html).

Puede significarnos una gran cantidad de documentos relacionados al diseño de hipertextos, y páginas web pero excluirá aquellos que utilicen html.

Podría decirse, que el proceso para convertir una necesidad de información en una ecuación de búsqueda es el siguiente:

FASE 1. Obtención de términos de búsqueda: Consisten en traducir una necesidad de información a conceptos unitarios y éstos a términos del lenguaje natural. Recomendamos, convertir la necesidad de información en una pregunta y extraer de ella los conceptos.

Ejemplo: Filmografía del realizador de Full Monty «Full Monty» AND «filmografía»

Esta búsqueda por estar en español puede arrojarle algún resultado útil en su búsqueda pero la mayoría de los resultados se orientarían hacia la palabra filmografía.

FASE 2. Obtención de sinónimos. Consiste en derivar sinónimos de cada concepto si es el caso.

Ejemplo: Puede utilizarse el nombre completo del director «peter cattaneo» AND «filmography»

Para así ir directamente a los recursos en inglés sobre este director de cine en particular.

En internet, lamentablemente todavía falta mucho para que el español tenga una presencia importante y podamos realizar nuestras búsquedas sólo en nuestro idioma. Por el momento, le recomendamos utilizar el inglés para las palabras claves y después traducir los resultados con los recursos ofrecidos por los motores de búsqueda.

FASE 3. Combinación: Consiste en combinar los conceptos y los sinónimos con los operadores adecuados y construir una ecuación formal.

Es importante que una vez se tengan diferentes opciones de búsqueda, combinarlas adecuadamente con and, or, not y las comillas para separar frases o nombres completos.

Se recomienda no utilizar mayúsculas en los nombres porque sólo hará búsquedas exactas.

Eventualmente:

FASE 4. Refinamiento: En ocasiones, el mismo buscador puede ayudarle a refinar su búsqueda, incluyendo o sacando algunos términos partiendo de los resultados de la primera búsqueda. La mayoría de los motores de búsqueda tienen este recurso y puede ser muy útil para dar con lo que estamos buscando.

EXPLOTANDO LOS YACIMIENTOS DE LA RED

«Una de las múltiples paradojas de la información, sin embargo, es que cuanto más útil es una fuente más difícil resulta separar de ella el ruido de la información. La causa es que una fuente es tanto más útil cuanto mayor sea la probabilidad de que contenga información relevante pero para incrementar esa posibilidad no hay más remedio que incrementar el número absoluto de documentos que contiene.», explica el profesor Lluís Codina, en su artículo «Búsqueda y obtención de información en internet»².

En resumen, a estas horas, resulta fundamental para un profesional aprender a manejar esas herramientas para que dejen de ser una fuente de ansiedad y lo lleven a convertirse en un experto tal como lo es en la actualidad con un índice, una hemeroteca o un simple fichero de biblioteca.

Entonces, deje de revolcarse en su propia autocompasión y supuesta incapacidad y póngase a trabajar, porque explotar los yacimientos de la red puede llegar a ser mucho más fácil de lo que piensa.

Por el momento, nos vamos a dedicar a analizar las herramientas que nosotros denominamos «sin barreras», es decir que o bien son gratuitas o bien no exigen el pago de una cuota regular, sino que pueden utilizarse libremente. Estas herramientas de búsqueda en internet las dividiremos en esta ocasión en Navegación y recuperación de información.

Navegación

Recuperación de Información

1. Directorios
2. Metadirectorios
3. Motores de búsqueda
4. Metabuscadores
5. Servicios de difusión selectiva de la información
6. Bases de datos especializadas

Es importante empezar por destacar la diferencia esencial que existe entre los directorios y los sistemas de recuperación de información, de los cuales los motores de búsqueda son los sistemas más conocidos. De hecho, hemos presentado la tabla anterior de manera que esa diferencia quede enfatizada.

NAVEGACIÓN VS RECUPERACIÓN

En primer lugar un directorio se diferencia de un motor de búsqueda en que clasifica los recursos en lugar de representarlos

mediante un conjunto de términos o de palabras clave. Un directorio, por tanto, consiste en una estructura jerárquica formada por clases y subclases. Cada una de estas clases contiene a un número determinado de recursos. El acceso a la información se realiza por navegación o desplazamientos sucesivos entre las clases y los niveles de la jerarquía.

Los recursos se seleccionan, analizan y clasifican de forma intelectual y por este motivo, solo contienen a una pequeña parte de la red. Los recursos clasificados por los directorios se cuentan por miles o decenas de miles, cuando los recursos o documentos publicados en internet se cuentan por decenas de millones.

El ejemplo clásico de un directorio es Yahoo! (www.yahoo.com), este directorio aparte de ser uno de los más grandes y el primero de ellos, también está actualmente esforzándose en sacar una versión para cada país y está dando sus primeros pasos en Latinoamérica.

Para localizar información en Yahoo, debe partirse de alguna de sus clases principales y procediendo por ensayo y error recorrer toda su estructura que es la forma más recomendable de utilización.

Su uso, es recomendable si comenzamos una actividad de búsqueda y queremos tener una visión global sobre un tema. También es sumamente recomendable cuando necesitamos encontrar recursos sobre un tema de una manera no específica.

La razón es que, Yahoo no recoge toda la información que hay en Internet, sino una parte más bien pequeña y además la manera de representar los documentos en Yahoo es muy sintética y puede no ser la más aconsejable.

MOTORES DE BÚSQUEDA

Los motores de búsqueda por su parte, proporcionan la consulta de los índices analíticos, los cuales representan el contenido de los documentos publicados en internet. Los motores de búsqueda generalista intentan representar la totalidad de la internet sobre la base de situar en sus índices todos y cada uno de los términos representativos contenidos en todos y cada uno de los documentos publicados en internet.

De este modo, cada recurso o documento se representa mediante un conjunto de palabras o frases, llamados términos de indexación porque forman parte del mencionado índice. Se supone que este conjunto de términos de indexación son

característicos del contenido del documento, de modo que si este artículo que el lector tiene en sus manos estuviera indexado por ese procedimiento, el conjunto de término que generaría sería el siguiente: Comunicación, Internet, directorios, motores de búsqueda, indexación, índices, documentos, recursos.

Normalmente un motor de búsqueda, compara los términos de la pregunta que formuló el usuario con los que figuran en el índice y selecciona de este modo todos los documentos o recursos que coinciden, totalmente o en parte con dicha expresión de búsqueda.

De allí que sea tan importante aprender a formular lo mejor posible nuestra ecuación de búsqueda para que el motor extraiga de su índice los documentos más acertados. Sin embargo, cuando se realiza una búsqueda a través de estos motores subyace una hipótesis que podríamos decir que es doble.

En el lado del motor de búsqueda, la hipótesis consiste en que el motor es capaz de extraer adecuadamente los términos que representan el contenido de los documentos. En el lado del usuario que plantea la pregunta, la hipótesis consiste en que los términos que utiliza estarán presentes en los documentos relevantes y no lo estarán en los documentos no relevantes.

Pero los hechos demuestran que solamente se cumplen parcialmente. «En primer lugar los motores de búsqueda son capaces de identificar cadenas de caracteres pero no conceptos, con lo cual toda vaguedad, ambigüedad, del lenguaje natural se traspaasa a los índices»³

Por lo tanto, el usuario deberá buscar otra manera de expresar su necesidad de información, y termine descubriendo a través del ensayo y error que algunos términos o sinónimos nada evidentes al principio sean los adecuados para esa búsqueda. Por ejemplo puede ser que al buscar arte impresionista sea más efectivo buscar por «Monet» que intentar por «arte» AND «impresionismo».

Uno de los motores de búsqueda más populares es Altavista (www.altavista.com) Esta empresa posee (supuestamente) el sistema de análisis de documentos más eficiente y por tanto es el que proporciona mayor número de respuestas relevantes y relacionadas con nuestras búsquedas.

Además cuenta con una cantidad de recursos bastante útil. Por ejemplo la opción «refine», muestra en forma gráfica o en forma de lista de términos, un análisis

de la pregunta. Podemos añadir esos términos a la búsqueda u obviarlos. Su eficacia es muy relativa así que no podemos asegurar que funciona a la perfección.

Otro recurso es el traductor de páginas que convierte la página encontrada al idioma que el usuario seleccione. El resultado puede ser más o menos comprensible dependiendo del estilo del lenguaje y lo complicado que sea el documento.

Difusión selectiva de la información

Este servicio no está actualmente muy generalizado porque está todavía en etapa de consolidación tecnológica. Sin embargo, constituye para los investigadores uno de los recursos más importantes.

Permite registrar un perfil temático de búsqueda a nombre de un usuario y a petición de este usuario periódicamente y de manera automática, actualizará su búsqueda localizando únicamente los últimos documentos referentes al perfil que se le ha indicado.

Uno de los más conocidos es el Scout Project (www.scout.cs.wisc.edu/scout) visítelo.

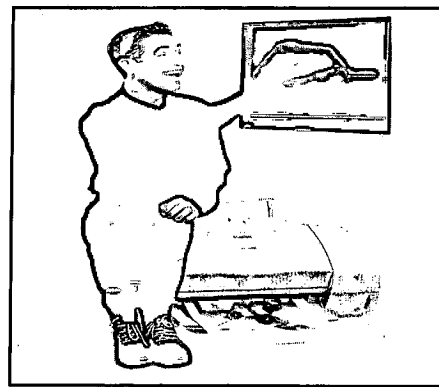
Motores como cerebros

Ese quizás sea el sueño de unos cuantos científicos, ingenieros o investigadores que probablemente estén en estos momentos encerrados en sus laboratorios tratando de sorprendernos con un frankenstein que solucione todos nuestros problemas y nos ayude a sobrepasar la ansiedad que esta sobresaturación de información nos está provocando a algunos.

Pero lo que si es cierto, es que a los motores de búsqueda (infinitamente útiles pero infinitamente falibles) todavía no pueden traducir a la perfección nuestras necesidades de información y no nos queda otra que adaptarnos a ellos.

Al ritmo que avanza Internet, parece que el problema no dejará de agravarse por lo que cada vez más, urge una solución a la forma en que los motores de indexación manejan actualmente la información. Esa solución se está investigando y ensayando en varios campos a la vez; en el procesamiento del lenguaje natural; en la incorporación de lenguajes documentales a los motores de búsqueda e indexación, en la aplicación de redes neuronales y finalmente, en la combinación de todos los procedimientos mencionados.

Consiste esta vía en ir configurando una especie de Internet II de facto que consiste en la creación de redes de tipo académico que analizan y describen los



recursos con el uso de meta datos.

Los meta datos, según explica Codina⁴, constituyen una forma de precisar de una manera muy rigurosa el contenido y otros atributos esenciales de un documento (como el responsable intelectual, la institución para la que trabaja, la lengua del documento, la relación con otros recursos, etc) que se incorporan al propio documento.

Este método podría ayudarnos mucho más a la hora de escoger una fuente confiable y un documento con base real y comprobada.

Todavía no está entre nosotros esta tecnología pero a poco se irá incorporando y tendremos que irnos adaptando a su nomenclatura. Lo importante es mantenerse al día, conectado e incorporando la tecnología a nuestro servicio, porque a fin de cuentas para eso está, no deje que la ansiedad y la sobresaturación de información lo subyugue, póngala a trabajar para usted y navegue aguas profundas como si estuviera en la comodidad de su hogar □

REFERENCIAS

- Impresos**
 ANDRIEU, O. *Cómo buscar y encontrar en internet*. Barcelona. Ediciones Gestión 2000, 1997, 191 pp.
 CODINA, Lluís. *Búsqueda y obtención de información en Internet*. Barcelona. IDEC, 1998. 27 pp
 HERNÁNDEZ MONTOYA, Roberto. *Revista Internet World*
- Recursos y Documentos digitales**
 Documento de ayuda en altavista. http://doc.altavista.com/help/search/search_help.shtml
 Ayuda en Yahoo en español
<http://es.docs.yahoo.com/info/ayuda.html>
- 1 HERNANDEZ MONTOYA, Roberto. «Periodismo. Informar será lo de menos». *Revista Internet World*. Pp 52-53. Número 3, Septiembre 1999.
 - 2 CODINA, Lluís. *Búsqueda y obtención de información en Internet*. Barcelona: IDEC, 1998, 27 pp (documento reprografiado)
 - 3 Idem
 - 4 Idem