

UNIVERSIDAD AUTÓNOMA METROPOLITANA. UNIDAD
XOCHIMILCO
UNIVERSIDAD DE LA HABANA. FACULTAD DE ECONOMÍA
UNIVERSIDAD DE MURCIA

PROGRAMA INTERNACIONAL DE MAESTRÍA EN GESTIÓN DE LA
INFORMACIÓN

Trabajo final del Seminario de Tecnologías de la Información y las
Comunicaciones.

Profesor: Dr. Ing. Armando de Jesús Plasencia Salgueiro

Motor de búsqueda especializado de la
FMVZ-UNAM

<http://www.fmvz.unam.mx>

<http://tics-gio.blogspot.com> (*)

Alumna: Nora Lucía Galván Ochoa

31 de julio de 2008

(*) URL alternativo, en caso de que las autoridades de la Facultad de Medicina Veterinaria y Zootecnia de la UNAM decidieran quitar el buscador para la fecha en que sea revisado el presente trabajo, **tics-gio** significa: **t**ecnologías de la **i**nformación y **c**omunicación - **g**estión de la **i**nformación en las **o**rganizaciones.

Resumen.

Motor de búsqueda especializado de la FMVZ-UNAM¹ es el trabajo final del *Seminario de Tecnologías de la Información y Comunicaciones* en la Maestría en Gestión de la Información. Se presenta el entorno de internet con un enfoque sobre el uso de las herramientas de búsqueda de información, la metodología llevada a cabo para el desarrollo de un buscador especializado instalado en la página de la FMVZ-UNAM (<http://www.fmvz.unam.mx>)², utilizando la tecnología desarrollada por Google (<http://www.google.es/coop/cse/>). Finalmente se presentan recomendaciones y consideraciones en el empleo de este tipo de herramientas.

Introducción.

De los servicios que Internet ofrece, el web es un enorme entorno de recuperación de información, en el que para publicar no es necesario apegarse a reglas en cuanto a estructura(1). A mediados de los años 90's Internet se ve transformado por la aparición de los motores de búsqueda y se estima que el 95% del tráfico de Internet se debe al uso de los motores de búsqueda y el 85% de los usuarios que tratan de recuperar información en el web lo hacen a través de los buscadores. Los buscadores son programas que consultan el contenido de los recursos en los sitios web y almacenan lo esencial de ese contenido en sus bases de datos, lo que da como resultado que las búsquedas que realizan los usuarios no sean eficientes. Por otro lado se han realizado estudios sobre el comportamiento de los usuarios ante las hojas de resultados que arrojan los buscadores, y se ha encontrado que el 90% de los usuarios consultan solo la primera hoja de resultados mientras que el 1% llega a consultar hasta la tercera hoja (2). La información electrónica se duplica cada 12 a 18 meses, por lo que localizar y acceder a la información en internet puede convertirse en un proceso un tanto complicado(3).

Es por ello que surge la necesidad de los usuarios en el empleo de buscadores especializados para poder acercarse a información específica que requieren.

En la actualidad los editores de información electrónica están muy interesados en que sus publicaciones sean indexadas por los buscadores más populares como es el caso de Google (<http://www.google.com>). Como empresa Google ofrece a los usuarios de internet más de 60 productos que en su mayoría son de acceso libre. Dentro de sus servicios Google ha desarrollado un motor de búsqueda con tecnología suficientemente consolidada y muy seguramente su popularidad se debe a la premisa de tomar en cuenta el comportamiento y necesidades de los usuarios para el desarrollo de sus productos es decir sus productos están enfocados al comportamiento del usuario. El buscador de Google trabaja en una manera

¹ FMVZ-UNAM son las siglas de Facultad de Medicina Veterinaria y Zootecnia de la Universidad Nacional Autónoma de México.

² <http://tics-gio.blogspot.com> es una página personal (blog) desarrollada como URL alternativo en caso de que las autoridades de la Facultad de Medicina Veterinaria y Zootecnia de la UNAM decidieran quitar el buscador para la fecha en que sea revisado el presente trabajo, **tics-gio** significa: **t**ecnologías de la **i**nformación y **c**omunicación - **g**estión de la **i**nformación en las **o**rganizaciones. Esto con la finalidad de mostrar el funcionamiento del buscador especializado y cubrir el requisito en el seminario.

masiva, con interfaz sencilla de utilizar y entrega en la mayoría de las veces resultados satisfactorios a los usuarios que lo manejan(3).

Las bases de datos de Google indexan documentos en diversos formatos que van desde procesadores de texto, hojas de cálculo, presentaciones, hojas html y formatos portables de diversos fabricantes. Como cualquier herramienta de cómputo el buscador de Google tiene también sus limitaciones, porque no puede buscar dentro de bases de datos de sitios, o en páginas generadas en forma dinámica o bien en aquéllas que solo se muestran cuando el usuario se registra para tener acceso (3),(4).

Paralelamente al crecimiento de la publicación electrónica en internet, se han desarrollado otros servicios como los buscadores especializados que permiten que los usuarios se acerquen pronta y eficazmente a información específica, un ejemplo de ello se puede desarrollar a través de Google (<http://www.google.es/coop/cse/>).

Esta herramienta se puede emplear en páginas personales o sitios completos de internet y está adaptada y creada con el núcleo de la tecnología de búsqueda de Google, para dar prioridad o restringir los resultados de búsqueda basados en los sitios web y en las páginas que se especifique(5).

Desarrollo

La metodología que se llevó a cabo para desarrollar el buscador personalizado de la FMVZ-UNAM (<http://www.fmvz.unam.mx>) utilizando el servicio proporcionado por Google se describe a continuación, sin embargo, el procedimiento se puede consultar en <http://www.google.es/coop/cse/>

1. Se creó una cuenta de usuario en el correo de Google: <http://mail.google.com>
2. Se activa la página de Google que ofrece el servicio de creación de motores de búsqueda personalizado <http://www.google.es/coop/cse/>, y se debe activar el botón de [Crear un motor de búsqueda personalizado].
3. Proporcionar nombre de usuario y contraseña del servicio de correo de Google (gmail).
4. Se inicia la sección de **Configurar su motor de búsqueda**, donde se debe proporcionar:
 - a. Información básica
 - i. Nombre del motor de búsqueda
 - ii. Descripción del motor de búsqueda
 - iii. Palabras clave del motor de búsqueda
 - iv. Idioma del motor de búsqueda. Se seleccionó Español.
 - b. ¿Qué quiere buscar?. En esta sección se debe especificar dónde va a trabajar el buscador personalizado, y se tiene las siguientes opciones:
 - Sólo los sitios que seleccione
 - En toda la web, pero enfatizar los sitios que selecciono
 - Toda la web

En el buscador personalizado de la FMVZ-UNAM (<http://www.fmvz.unam.mx>) se seleccionó la primera opción (Sólo sitios que seleccione) con la finalidad de restringir la búsqueda a los sitios de la Facultad.

- c. Seleccione algunos sitios: En esta sección se proporcionó la siguiente configuración: www.fmvz.unam.mx/* para incluir todas las páginas de la FMVZ-UNAM.
- d. Edición. Se proporcionan dos tipos de ediciones:

- ◉ **Edición estándar**: es gratuita, y las páginas de resultados deben mostrar anuncios.
 - No mostrar publicidad en las páginas de resultados (sólo para sitios sin ánimo de lucro, universidades y agencias gubernamentales).
- **Edición para empresas**: desde 100 euros al año. No aparecen anuncios en las páginas de resultados.

Se seleccionó la Edición estándar y que no se muestre publicidad, debido a que es un sitio académico.

- e. Se activa el botón de [Siguiete]
- f. Se habilita la opción de que se envíe el código del buscador a la cuenta de correo de Google.
- g. El código que se recibe se debe colocar en la sección de la página donde se desea mostrar. El siguiente es el código que se recibe:

```
<form action="http://www.google.com/cse" id="cse-search-box">
  <div>
    <input type="hidden" name="cx" value="010234395181439621419:mzrtclvnzxcg" />
    <input type="hidden" name="ie" value="UTF-8" />
    <input type="text" name="q" size="31" />
    <input type="submit" name="sa" value="Buscar" />
  </div>
</form>
<script type="text/javascript" src="http://www.google.com/coop/cse/brand?form=cse-search-box&lang=es"></script>
```

NOTA ACLARATORIA –IMPORTANTE-

Para la instalación del buscador personalizado en la FMVZ-UNAM, se pidió autorización al jefe del departamento de cómputo el Ing. Miguel Álvaro Pérez Hernández, quien es el administrador de la red de esta dependencia y a la vez jefe de la autora del presente trabajo. Se espera que el buscador quede funcionando permanentemente, en caso contrario se creó una página personal (blog) <http://tics-gio.blogspot.com> como URL alternativo para poder demostrar que si se llevó a cabo el desarrollo del buscador personalizado, y en el cual se aplica la metodología anteriormente descrita.

Recomendaciones

El tema central de la tesis que la autora está desarrollando es sobre la implementación y evaluación de metadatos en los recursos electrónicos disponible en internet. A través de lo que se ha estudiado se ha encontrado que es muy recomendable que las publicaciones electrónicas contengan metadatos con especial énfasis el campo de título, además que éste sea el adecuado para describir claramente cada recurso electrónico publicado independientemente de la norma de metadatos que se utilice ya que se ha visto que se mejora significativamente la visibilidad de los recursos electrónicos ante los buscadores como Google. Además que una buena catalogación y diseño sencillo en los sitios web hacen más visible los recursos electrónicos ante los buscadores(4;6).

Consideraciones

El buscador personalizado de la Facultad de Medicina Veterinaria y Zootecnia de la UNAM quedó instalado en su página (<http://www.fmvz.unam.mx>), por lo que es importante considerar cómo hacer que los usuarios de internet en especial, la comunidad veterinaria realmente visite la página para poder emplear la herramienta de búsqueda desarrollada en este trabajo es decir se debe considerar la estrategia de promoción del sitio de la FMVZ-UNAM.

Otra consideración a tomar en cuenta es que conforme va creciendo el volumen de las colecciones de recursos electrónicos en un sitio web va aumentando los costos para que éstos queden adecuadamente indexados y catalogados por lo que es importante tener en cuenta la implementación de dicha metodología.

Referencias

- (1) López Carreño R. Taller. Recuperación y Gestión de la Información Académica en Internet. El Entorno Internet. 10-1-2007. Universidad Autónoma Metropolitana, Xochimilco, México, D.F., Programa de Superación Académica. Tronco Interdivisional.
Ref Type: Hearing
- (2) Zhang J, Dimitroff A. The impact of webpage content characteristics on webpage visibility in search engine results (Part I). *Information Processing and Management* 2004; 41(2005):665-690.
- (3) Arnold SE. Search: the new application platform. *Viewpoint. The Electronic Library* 2006; 24(2):121-125.
- (4) Dawson A, Hamilton V. Optimising metadata to make high-value content more accessible to Google users. *Journal of Documentation* 2006; 62(3):307-327.
- (5) Martínez Majado A. Álvaro Martínez Majado. Página personal. 2008.
Ref Type: Personal Communication
- (6) Dawson A. Digital directions. Creating metadata that work for digital libraries and Google. *Library Review* 2004; 53(7):347-350.