

Cuil, ¿la competencia de *Google*?: una evaluación desde el punto de vista documental

Cuil, will it compete with Google?: an evaluation from the documentary point of view

Yudisbel Rojas Cruz¹; Asdrúbal Ramírez Hernández¹

¹Licenciada en Bibliotecología y Ciencias de la Información. Grupo de Arquitectura y Estándares de Información. Dirección Técnica de la Infraestructura Productiva. Universidad de las Ciencias Informáticas. Cuba.

¹Licenciada en Bibliotecología y Ciencias de la Información. Departamento de Ingeniería y Gestión de Software. Universidad de las Ciencias Informáticas. Cuba.

RESUMEN

Se realizó una evaluación cualitativa de los buscadores *Cuil* y *Google* sobre la base de los siguientes criterios de evaluación: búsqueda, recuperación, calidad de los contenidos, usabilidad y servicios adicionales. *Google* será el buscador más popular de la Web o al menos no será superado por *Cuil* en poco tiempo.

Palabras clave: Buscadores, evaluación, recuperación de la información.

ABSTRACT

A qualitative evaluation of Cuil and Google search engines was made on the basis of the following evaluation criteria: search, retrieval, content quality, usability and additional services. Google will be the most popular search engine of the Web, or at least, it won't be exceeded by Google in a short time.

Key words: Search engines, evaluation, information retrieval.

Paralelamente al aumento de los contenidos en Internet se desarrollan disímiles buscadores para recuperar información en la Web. Los primeros consistían en directorios de páginas organizados manualmente; sin embargo, no fueron suficientes, y fue humanamente imposible lograr que pudieran identificar, registrar e indizar el volumen —cada vez mayor— de los contenidos que aparecían a un ritmo cada vez más acelerado. Es así como aparecen los motores de indización automáticos que, si bien resolvieron el problema hasta cierto punto, también generaron otros no menos importantes, como la inconsistencia en la recuperación de los resultados y su falta de relevancia.

En este contexto surgieron iniciativas para facilitar un marco objetivo para la evaluación de los sistemas de recuperación de información, como la serie de conferencias denominadas *Text REtrieval Conference* (TREC), dedicadas a examinar diversos aspectos en esta área del conocimiento; *Evaluation Forum* (CLEF); NII-NACSIS Test Collection for IR Systems (NTCIR) y la Conferencia de Aprendizaje sobre el Lenguaje Natural (CONIL).

Se pretende en esta ocasión evaluar, desde un punto de vista documental, dos buscadores Web objeto actualmente de una controvertida discusión en Internet: *Google* (www.google.com) y *Cuil* (www.cuil.com).

MÉTODOS

A partir de la opinión generalizada en Internet sobre la competencia que puede representar *Cuil* para *Google*, este último considerado por muchos expertos como el mayor buscador de la red, se decidió comparar, según criterios, ambos buscadores.

Elaboración de los criterios de comparación: Tras un análisis sobre el estado del arte en materia de evaluación de buscadores en Internet, se decidió dividir los criterios existentes en cinco grupos: búsqueda, recuperación, calidad de los contenidos, usabilidad y servicios adicionales.

Evaluación de los criterios seleccionados: Se realizó un grupo de consultas a expertos para determinar la validez de los criterios escogidos. Cada uno de los criterios se detalla a continuación.

1. *Búsqueda:* "Un sistema de búsqueda es mejor cuanto más flexible es y cuantas más posibilidades de recuperación ofrece".¹ Se elaboraron diversas estrategias de búsqueda. Se realizaron consultas con diferentes estrategias:

- Con una única palabra no ambigua: *Informática*.
- Con varias palabras: *Informática en la salud*.
- Con varias palabras, de las cuales al menos una resulta ambigua pero tiene sentido con respecto al resto de la consulta. Por ejemplo, *cáncer* puede referirse a uno de los signos del zodiaco o al nombre de una enfermedad. En este

caso, los resultados dependen del resto de las palabras de la búsqueda: tratamiento del cáncer, tipos de cáncer, etcétera.

2. *Recuperación*: Se refiere a la forma de presentación de los documentos que componen un sistema documental. En relación con el orden de presentación de los resultados, es importante que sea posible seleccionar entre varios criterios de ordenamiento. En el caso de los motores de búsqueda, al gestionar los URL es recomendable que presenten las páginas recuperadas agrupadas por servidores.

En ambos buscadores se ejecutaron las estrategias de búsquedas elaboradas y se evaluaron individualmente cada uno de los resultados.

3. *Calidad de los contenidos*: Se refiere fundamentalmente a la relevancia de los documentos obtenidos en la primera página de resultados.

4. *Usabilidad*: "...la capacidad de un software de ser comprendido, aprendido, utilizado y resultar atractivo para el usuario, en condiciones específicas de uso";² "...la eficiencia y satisfacción con la que un producto permite alcanzar objetivos determinados a usuarios específicos en un contexto de uso específico".³ Este criterio se evaluó mediante el método de inspección o evaluación heurística, que toma como referencia los principios heurísticos de *Jakob Nielsen*.^{4,5} La revisión se realizó de manera individual por dos evaluadores que asumieron el papel de usuarios.

5. *Servicios adicionales*: Funcionalismos que agregan valor a la búsqueda y recuperación de la información, como son las herramientas de idioma, *bookmark*, *blog*, etcétera.

RESULTADOS Y DISCUSIÓN

A continuación se muestra un cuadro comparativo entre los buscadores *Cuill* y *Google*:

Criterio	Descripción	Cuill	Google
Búsqueda	Formularios de búsqueda	No ofrece la posibilidad de elegir entre un formulario simple y otro más detallado. Por el momento el formulario sólo permite búsquedas simples.	Presenta un buscador simple y uno avanzado con las opciones de: mostrar resultados, idiomas, región, formato de archivo, fecha, presencia, dominios, derechos de uso, filtrado, páginas similares y enlaces relativos. Además, permite buscar por tipos de contenidos: noticias, imágenes y en la Web.
	Herramientas de	Uso de operadores	Accepta la utilización de

	búsqueda	<p>lógicos: reconoce solo el AND y el +. Para otros como el OR y el NOT, devuelve prácticamente los mismos resultados que para el AND.</p> <p>Operadores de posición: El NEAR, aunque devuelve resultados, no son muy relevantes porque existe cierta lejanía de los términos, incluso en los primeros resultados. El BEFORE es un poco mejor reconocido. No se notan grandes cambios entre el uso del NEAR y el FAR. Operadores de exactitud y truncado: no reconoce el * y el ?. El () aunque lo reconoce, no ofrece resultados relevantes; en cambio, si se utilizan las "", la relevancia es mayor.</p>	<p>los operadores lógicos. AND recupera sólo las páginas que incluyen todos los términos de búsqueda y utiliza:</p> <p>- para excluir un término.</p> <p>"" para buscar frases.</p> <p>+ para incluir palabras vacías en la búsqueda.</p>
	Campos de búsqueda	<p>No es posible dirigir las búsquedas a campos determinados como: título, URL y descripción.</p>	<p>Permite comandos de búsquedas como:</p> <p><i>link</i>: para buscar páginas que tienen vínculos hacia el URL especificado.</p> <p><i>inurl</i>: para buscar páginas que tienen el primer término en algún lugar del URL.</p> <p><i>allinurl</i>: para buscar páginas que tienen todos los términos en algún lugar del URL.</p> <p><i>intitle</i>: para buscar páginas que tienen el primer término en algún lugar del título.</p> <p><i>allintitle</i>: para buscar</p>

			archivos (12 formatos de archivos), <i>site</i> : para restringir la búsqueda a un sitio indicado. <i>related</i> : para buscar otras páginas con patrones de enlaces similares a los del URL especificado y con niveles de jerarquía similares, <i>define</i> : permite ver la definición de una palabra o frase. ~: se utiliza para buscar considerando sinónimos (sólo para la Web y en inglés).
	Control del vocabulario	No posee sistema para controlar el vocabulario, pero presenta una sugerencia de términos a buscar.	No posee un sistema de control del vocabulario aunque sugiere luego de la búsqueda un conjunto de temáticas relacionadas con la de la búsqueda, así como sugerencias de ortografía a la palabra clave introducida.
	Detección de novedades	No las delimita.	No las delimita.
Recuperación	Relevancia de la primera página de resultados	No siempre devuelve documentos relevantes en función de la búsqueda realizada.	Ofrece registros relevantes en la primera página e, incluso, en otras páginas.
	Tiempo de respuesta	Rápido	Rápido
	Interfaz	Posee una interfaz muy bien estructurada por secciones y campos, ofrece una división espacial muy bien definida por contenidos a lo que se le suman imágenes por resultados de búsqueda.	Simple y directa, se ofrece de manera muy básica.
Calidad de contenido	Cobertura	Posee mayor cobertura, se habla de que posee el triple de páginas indizadas que <i>Google</i> .	Cobertura especializada en distintos tipo de documentos: (.pdf, .ps, .doc, .xls, .txt, .ppt, .rtf, .asp, .wpd, .jpg y otros)
	Actualización de	No se conoce.	Mensual, durante 4 o 6

	las bases de datos		días.
	Procesamiento de la información	El buscador trabaja muy directamente con los contenidos del sitio, se centra en el procesamiento de estos, una presentación de resúmenes por resultados y su categorización.	Posee un procesamiento basado en el <i>Page Rank</i> , algoritmo para el posicionamiento de las páginas que no se conoce a profundidad, pero que no sólo beneficia las visitas a la página sino también otras aristas.
	Funcionamiento de las conexiones	Buen funcionamiento.	No se encontraron enlaces rotos en las tres primeras páginas de resultados.
	Número de páginas cubiertas	121 617 892 992 páginas Web	Comprende más de 8 000 millones de direcciones URL.
Usabilidad	Navegación	Posee una navegación por temáticas o categorías con dos subsistemas de navegación arriba y a la derecha en este sentido. Además, presenta una navegación estándar por resultados.	Posee una navegación sencilla y asequible, donde se delimitan los resultados por tipos de formatos.
	Funcionalidad	Poco trabajo con otros idiomas, sólo trabaja a profundidad con el inglés.	Políglota, posee interfaces y realiza búsquedas en 34 idiomas. De esto se infiere la profundidad de sus funciones.
	Control del usuario	Es algo complejo debido a las funcionalidades que ofrece, en el sentido de las posiciones de la información y las categorías predeterminadas que no pueden modificarse por el usuario.	En el servicio <i>igoogle</i> , el usuario utiliza bloques de información elaborados con tecnología AJAX, y puede personalizar su propia página de <i>Google</i> .
	Lenguaje y contenido	Posee un lenguaje especializado en sus subcategorías.	Lenguaje abierto y coloquial.
	Ayuda en línea	Posee ayuda, aunque es pobre, y no corrige los errores introducidos a la hora de realizar las	Posee una ayuda amplia que, además de la búsqueda, abarca el uso de cada uno de los

		búsquedas.	servicios.
	Información del sistema	Alta disponibilidad.	Alta disponibilidad.
	Accesibilidad	Es accesible con todo tipo de hardware.	Es accesible con todo tipo de hardware.
	Coherencia	Existe coherencia entre los aspectos formales del buscador como colores, tipografías e imágenes.	Existe coherencia entre los aspectos formales del buscador como colores, tipografías e imágenes.
	Prevención errores	Baja prevención de errores.	Posee un servicio de prevención de errores ortográficos.
	Claridad arquitectónica	Se hace un tanto compleja para usuarios poco relacionados con los buscadores y la Web en general.	Sencilla y asequible.
Servicios adicionales		Presenta un filtro de sitios no deseados, como los pornográficos.	Traductor que soporta varios idiomas, <i>Titulares de noticias, calculadora, conversión de moneda, traducción de páginas Web, páginas similares</i> , entre otros.

CONCLUSIONES

Cuil ha tratado de promocionarse en el mercado sobre la base de los puntos débiles de *Google* como la privacidad de los usuarios y la manera de posicionar los contenidos, y asumió una filosofía anti *Page Rank* que, sin embargo, no logra aún satisfacer las necesidades informativas de los usuarios con los resultados que devuelve. Propone una nueva interfaz muy atractiva al usuario para organizar los resultados, y asume los principios del ahorro de energía.

Google presenta un avance superlativo en relación con la posición en el mercado, trabajo con idiomas y otros servicios que le agregan valor; sin embargo, *Cuil* promete una aún no lograda mejor estructura de contenidos.

La calidad de los contenidos recuperados sigue siendo un problema presente en los dos buscadores, más allá de la capacidad que cada uno tiene para indizar un mayor número de páginas o la relevancia presente en los primeros resultados. Ambos poseen grandes bases de datos. En el caso de *Cuil* se desconoce la manera en que indiza las páginas. Las posibilidades de búsqueda actuales de *Google* son muy superiores al ofrecer mayor cantidad de servicios en este sentido. El tiempo de respuesta en la recuperación es rápido en los dos buscadores, cuestión en la cual han centrado su trabajo.

Finalmente, se considera que *Google* seguirá siendo el buscador más popular de la Web o al menos este no será superado por *Cuil* en poco tiempo.

REFERENCIAS BIBLIOGRÁFICAS

1. Maldonado A, Fernández E. Evaluación de los principales buscadores desde un punto de vista documental: recogida, análisis y recuperación de recursos de información. 1988. Disponible en: http://fesabid98.florida-uni.es/Comunicaciones/a_maldonado/A_Maldonado.htm [Consultado: 27 de septiembre de 2008].
2. ISO. Ergonomic requirements for office work with visual display terminal. Norma ISO/FDIS 9241-11. Genova: ISO. 1998.
3. _____. ISO 9126 Standard. Disponible en: <http://www.issco.unige.ch/ewg95/node14.html#SECTION00311000000000000000> [Consultado: 27 de septiembre de 2008].
4. Nielsen J. How to conduct a heuristic evaluation. Disponible en: http://www.useit.com/papers/heuristic/heuristic_evaluation.html [Consultado: 27 de septiembre de 2008].
5. _____. Ten usability heuristics. 2002. Disponible en: http://www.useit.com/papers/heuristic/heuristic_list.html [Consultado: 27 de septiembre de 2008].

Recibido: 4 de noviembre de 2008.

Aprobado: 14 de noviembre de 2008.

Lic. *Yudisbel Rojas Cruz*. Universidad de las Ciencias Informáticas. Carretera a San Antonio Km 2 ½ Rpto. Torrens, Boyeros. Ciudad de La Habana. Cuba. Correo electrónico: yudisbel@uci.cu

Ficha de procesamiento

Términos sugeridos para la indización

Según DeCS¹

ALMACENAMIENTO Y RECUPERACIÓN DE LA INFORMACIÓN; INTERNET.
INFORMATION STORAGE AND RETRIEVAL; INTERNET.

Según DeCI²

RECUPERACIÓN DE LA INFORMACIÓN ; INTERNET; MOTORES DE BUSQUEDA;
METABUSCADORES.
INFORMATION RETRIEVAL; INTERNET; SEARCH ENGINE; META-SEARCH ENGINE.

¹BIREME. Descriptores en Ciencias de la Salud (DeCS). Sao Paulo: BIREME, 2004.
Disponible en: <http://decs.bvs.br/E/homepagee.htm>

²Díaz del Campo S. Propuesta de términos para la indización en Ciencias de la Información. Descriptores en Ciencias de la Información (DeCI). Disponible en: <http://cis.sld.cu/E/tesauro.pdf>

Copyright: © ECIMED. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 2.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente el autor o los autores y su fuente original.

Cita (Vancouver): Rojas Cruz Y, Ramírez Hernández A. Cuil, ¿la competencia de Google?: una evaluación desde el punto de vista documental. Acimed. 2008; 18(6). Disponible en: Dirección electrónica de la contribución [consultado: día/mes/año].