

RECUPERACIÓN DE INFORMACIÓN CENTRADA EN EL USUARIO Y SEO: CATEGORIZACIÓN Y DETERMINACIÓN DE LAS INTENCIONES DE BÚSQUEDA EN LA WEB*

LATEST TRENDS IN WEB INFORMATION RETRIEVAL
AND IN SEO FACTORS

CARLOS GONZALO, LLUÍS CODINA Y CRISTÒFOL ROVIRA
cgonzalop@yahoo.com | lluis.codina@upf.edu | cristofol.rovira@upf.edu

Universitat Pompeu Fabra
Grupo de Investigación DigiDoc

Resumen: Presentación de las últimas tendencias en recuperación de información en la web y en los factores de posicionamiento (SEO), cada vez más centrados en señales procedentes de los usuarios como: perfil de quien realiza las búsquedas e interpretación de la intención de búsqueda. El objetivo de los buscadores es doble: centrarse al máximo en los usuarios y hacer cada vez menos predecible la composición de la página de resultados (SERP) para luchar contra el *spam*. **Palabras clave:** búsqueda en la web; recuperación de información; posicionamiento Web; SEO; señales de usuario; página de resultados; SERP; *spam* en buscadores; Google.

Abstract: Latest trends in web information retrieval and in SEO factors, increasingly focused on signals from users as: profile of who performs the search and the interpretation of user intent. The objective of search engines is twofold: focusing at the maximum in the users and make ever less predictable the composition of the search engine result page (SERP), and combating spam. **Keywords:** Web search; information retrieval; SEO; users signals; search engine results page; SERP; search spam; Google.

[*] Este trabajo forma parte del proyecto 'Audiencias activas y periodismo. Interactividad, integración en la web y buscabilidad de la información periodística'. CS02012-39518-C04-02. Plan Nacional de I+D+i, Ministerio de Economía y Competitividad (España).

1. El factor usuario en la recuperación de información

Una de las ambiciones de la Recuperación de Información (RI) clásica (la anterior a la era Google) era la de incorporar los perfiles de los usuarios en las respuestas. En los años ochenta y noventa cualquier revisión del estado de la cuestión en ese ámbito quedaba incompleta sin la lista de las mejoras que en aquella época se echaba a faltar en la RI y uno de los aspectos más señalados era éste, precisamente. Entonces no lo sabíamos, pero tuvieron que pasar, entre otras, las siguientes cosas para que este objetivo empezara a ser resuelto:

- ▶ La expansión de la World Wide Web.
- ▶ El predominio de Google en la búsqueda.
- ▶ La necesidad de combatir el *spam*.

Las tres cosas anteriores se relacionan de la siguiente manera: con la expansión de la Web se hicieron cada vez más necesarios los motores de búsqueda; y entre ellos Google obtuvo el dominio que conocemos actualmente, gracias a su sistema de ordenación de resultados basado en el análisis de enlaces.

Pero en la primera década del siglo XXI su modelo de negocio estaba cada vez más amenazado por el cada vez más habilidoso *spam* en buscadores. Si su sistema de análisis de enlaces no era capaz ya de eliminar el *spam* (páginas fraudulentas o de baja calidad) en sus resultados de búsqueda, todo su negocio basado en publicidad estaba en peligro: los usuarios dejarían de confiar en un buscador cuya página de resultados no limitaba los primeros resultados a las mejores páginas.

En este punto, a partir de los sucesivos cambios en los algoritmos de búsqueda y ordenación que Google implementó entre 2012 y el momento actual (mediados del 2015), los resultados aparecidos en sus respuestas fueron siendo cada vez menos predecibles y, por tanto, menos manipulables.

Entre estos cambios hubo dos que se centraron en el usuario: uno de ellos en su perfil (procedencia, edad, intereses, etc.) y el otro en la intención de búsqueda expresada de forma directa o indirecta. Para poder entender todo esto es necesario ante todo recordar que Google se expresa a partir de lo que se denomina la Search Engine Results Page (SERP). En lo que sigue intentamos ofrecer un estado de la cuestión sobre cómo se generan actualmente las páginas de resultados que puede ser útil para los responsables de SEO de medios de comunicación y, por supuesto, de cualquier sitio intensivo en contenidos.

2. Composición de la página de resultados (SERP)

Aunque los buscadores pueden llegar a ofrecer millones de resultados para una sola búsqueda, en realidad esto carece de importancia. Desde el punto de vista del usuario final es indiferente que, para una palabra clave, el buscador presente ya sea unas pocas decenas o varios millones de resultados, ya que solamente examinará los primeros.

Así que, para cualquier búsqueda, los resultados que importan son los que aparecen en las primeras posiciones de la página de resultados o SERP (*Search Engine Results Page*). Por este motivo, conocer lo mejor posible cómo se generan SERP resulta de interés para los profesionales del SEO.

Lo más importante de la **SERP** es que su composición es el resultado de la combinación de, al menos, tres **grupos** de factores, cada uno de ellos, por sí solos, merecedores de atención. Se trata de los siguientes:

► Los **factores *on page* y *off page*** de cada una de las páginas filtradas por la **palabra clave**. Lógicamente, el primer factor es la palabra clave utilizada en la consulta, ya que es el disparador de todo el proceso y la que actúa como primer filtro. Solamente las páginas que posean esa palabra clave (según los datos del índice del buscador) pasarán a ser elegibles para formar parte de la SERP. Buena parte del SEO se centra en estos aspectos cuando estudia temas referentes al contenido y a los factores *on page*, pero también a los factores *off page* en aspectos como el texto de anclaje de los enlaces de entrada. En este apartado, los sitios web de medios de comunicación tienen de entrada una ventaja directa porque forma parte de su naturaleza crear sitios ricos en contenidos.

► El **perfil del usuario**. Como hemos señalado, desde hace algún tiempo, los buscadores —y muy especialmente Google— tienen en cuenta los parámetros más importantes del así llamado “perfil del usuario”. El primer componente es su ubicación geográfica, cosa que determina mediante la IP de su conexión y, en el caso de las búsquedas móviles, por geolocalización. Además, tiene presente su actividad anterior a través de las *cookies* instaladas y otros aspectos de sus intereses cuando es un usuario que ha iniciado sesión y, por tanto, se ha identificado. Aquí Google puede utilizar una amplia variedad de ‘señales’ como sus búsquedas anteriores, así como su perfil demográfico y profesional gracias a datos procedentes de otras cuentas, como Google+ o Gmail.

► La **intención de búsqueda** (*user intent*). Este aspecto es relativamente nuevo en el SEO, aunque existen estudios desde hace años. Su importancia parece cada vez mayor desde los algoritmos de Google introducidos a partir del 2012.

Por este motivo, dada su creciente importancia en el SEO, en este trabajo nos centraremos en el tercero de ellos, es decir, en el impacto que tiene en la SERP la intención de búsqueda, concretamente en el caso del buscador Google.

3. La intención de búsqueda

A través de estudios de usuario divulgados por Google, quedó bien establecido que las motivaciones de los usuarios para hacer consultas en buscadores suelen dividirse en tres grandes grupos:

- Informacionales.
- Transaccionales.
- Navegacionales.

Las búsquedas o palabras clave informacionales son las consultas que buscan obtener informaciones sobre algún tema y, en concreto, pueden buscar imágenes, canciones, vídeos o documentos. Las transaccionales son consultas que tienen una motivación comercial (típicamente, adquirir un producto o servicio específico). Por último, las consultas navegacionales son aquellas en las cuales lo que los usuarios desean es poder ir a una URL específica. Ahora bien, hasta ahora, esta clasificación resultaba de un ‘grano’ excesivamente grueso para poder operar con ella más allá de su simple enunciación.

Afortunadamente, estudios posteriores (como el de Jansen *et al.*, 2008) aportaron más información sobre este aspecto. En lo que sigue, presentaremos una taxonomía sobre las intenciones de búsqueda derivada de estos estudios, y que parten también de la triple caracterización ya enunciada.

3.1. Consultas con posible intención navegacional

Las pistas o señales que, según los analistas, utiliza Google para identificar esta clase de consultas son las siguientes:

- Las consultas que contienen:
 - ▷ Nombres de compañías, empresas, organizaciones, marcas o personas.

- ▷ El término “web” como parte de las palabras clave.
- ▶ Las consultas que presentan una longitud inferior a tres términos.

3.2. Consultas con posible intención transaccional

Por su parte, para las de tipo transaccional, los indicadores que indican esta intención son los siguientes:

- ▶ Las consultas que contienen términos relacionados con:
 - ▷ Películas, canciones, nombres de recetas.
 - ▷ La idea de ‘obtención’ de alguna cosa (por ejemplo, letras de canciones, recetas, etc.).
 - ▷ Verbos como ‘descargar’ seguido de términos vinculados con *software*, películas, series, música, aplicaciones, etc.
 - ▷ Audio, imágenes o vídeo.
 - ▷ Interacciones. Por ejemplo: comprar, vender, alquilar, etc.
- ▶ Las consultas que utilizan expresiones directas como ‘películas de [actor/director]’, ‘canciones de [grupo/cantante]’, ‘letras de [grupo/cantante]’, ‘imágenes de’, ‘multimedia’ o palabras clave compuestas por extensiones de archivos o de compresión (.jpeg, .zip, etc.).

3.3. Consultas con posible intención informacional

Por último, las señales presentes en la interrogación del usuario que sugieren una intención informacional son las siguientes:

- ▶ Consultas que incluyen signos o palabras de interrogación o términos que sugieren preguntas (por ejemplo, “maneras de”, “cómo”, “qué es”, etc.).
- ▶ Las que utilizan frases del lenguaje natural en lugar de ecuaciones formales o simples palabras clave yuxtapuestas.
- ▶ Las que contienen términos que denotan aspectos de organización o presentación de *reports* o informes (por ejemplo, lista de reproducción, etc.).

- ▶ Operaciones de búsquedas que consisten en un refinamiento de los términos de búsqueda de una consulta anterior.
- ▶ Operaciones de búsqueda efectuadas después de que el usuario haya visualizado múltiples instancias de una SERP.
- ▶ Consultas con más de dos términos.
- ▶ Por exclusión: consultas que no cumplen los criterios navegacionales o transaccionales.

Esta tipología básica de búsquedas aún se puede hacer más detallada o sofisticada. De hecho, Jansen *et al.* (2008) aplican una clasificación jerárquica de tres niveles basada en las características de cada tipo de consulta y que sintetiza los estudios previos.

El cuadro resultante de la taxonomía jerárquica de las intenciones de los usuarios expresadas en búsquedas es el siguiente:

Tabla 1. Clasificación de la intención del usuario

Nivel 01	Informacional				Navegacional		Transaccional				
Nivel 02	Directas	No directas	Lugar físico	Listas	Consejos	Navegación con objetivo informacional	Navegación con objetivo informacional	Obtener	Descargar	SERP	Interactuar
Nivel 03	Abierta	Cerrada						Off-line gratis	No Enlaces gratis	Otras	

Fuente: Jansen *et al.* (2008)

Para interpretar la tabla anterior, recordemos que en el **nivel 1** encontramos la división principal entre informacionales, navegacionales o transaccionales. En el **nivel 2**, las búsquedas informacionales se dividen entre directas (cuando hay una pregunta específica) y no directas (cuando son generalistas), o cuando el objetivo es encontrar un lugar del mundo físico, o conseguir listas de ítems o asesoramiento de algún tipo.

A su vez, las búsquedas navegacionales se clasifican según si su objetivo es transaccional de forma directa o si el objetivo es obtener información previa a una posible transacción. Además, las búsquedas transaccionales pueden ser para ‘obtener’, para ‘descargar’ o ‘interactuar’ o bien aquellas que tienen como objetivo final la propia SERP al dar solución a un problema de tipo informacional.

En el **nivel 3** las navegacionales no tienen división. Por su parte, la división de informacionales consiste en que puede ser abierta o cerrada según si buscan una respuesta concreta o amplia, y las transaccionales pueden ser del tipo *off-line*, gratis, no gratis, enlaces u otras.

4. Intención de búsqueda y preferencias de clics en SERP

No siempre los 10 primeros resultados de una SERP son visibles de forma directa ya que tanto la publicidad como los servicios de los buscadores desplazan los resultados orgánicos del área visible, llegando al extremo que en ocasiones sólo dos resultados del Top10 algorítmico aparecen en la zona visible.

Los autores Chris Barry y Mark Lardner (2011) muestran cuál es el comportamiento de los usuarios de buscadores en lo relativo a su interacción con las SERP y el primer *click* que realizan, constatando que las preferencias de *click* dependen de la tipología de búsqueda.

Así, en las búsquedas informacionales el primer *click* de la gran mayoría de usuarios es para la primera posición de la SERP, que recibe cinco más primeros *clicks* que el siguiente enlace orgánico. Respecto a las búsquedas transaccionales, los estudios atestiguan que el 95 por ciento de todos los primeros *clicks* ocurre entre el primer enlace patrocinado y el tercer enlace orgánico, todos ellos dentro del área visible de la SERP.

El patrón general de la valoración de los resultados orgánicos frente a los enlaces patrocinados es que en búsquedas informacionales los usuarios son más propensos a hacer *click* en enlaces orgánicos, en lugar de enlaces patrocinados. Los enlaces orgánicos son percibidos como más relevantes, incluso si los enlaces patrocinados se presentan como igual de relevantes que los orgánicos.

La percepción del usuario sobre los enlaces patrocinados y enlaces orgánicos en consultas transaccionales es que en búsquedas relacionadas con comercio electrónico los enlaces patrocinados son probablemente más relevantes que los vínculos orgánicos, y los enlaces patrocinados visualizados en la parte superior de la SERP son más relevantes que los enlaces patrocinados del lado derecho.

5. El 'Click-through' (CTR)

Este trabajo no estaría completo sin mencionar otro *input* procedente de los usuarios. Como ya sabemos, una de las fuentes más importantes de tráfico hacia un sitio web es Google a través de las búsquedas de los usuarios y los *clicks* que hacen en las páginas de resultados. De esta interacción se deriva el concepto de *click-through* (CTR), una métrica que calcula la cantidad de veces que un enlace ha sido clicado en relación a la cantidad de veces que ha sido visible.

Esta métrica también es el principal indicador de rendimiento o KPI (*Key Performance Indicator*) del tráfico generado a través desde las SERP de los buscadores y uno de los factores que más influyen en el posicionamiento. Por ello, los analistas estiman que también es uno de los que más utilizan los propios buscadores para la reconfiguración de sus SERP en base a la interacción con los usuarios. Se considera que Google calcula el CTR global de la página así como el CTR para búsquedas específicas e introduce cambios en la posición del Top10 en base a la interacción de los usuarios.

5. Conclusiones

Lo primero que podemos concluir es que Google intenta hacer cada vez menos predecibles la composición de sus SERP. El motivo, a su vez, es contrarrestar tanto el *blackhat* SEO como la mera sobre-optimización SEO.

Al construir sus SERP como una combinación de factores *on page*, factores *off page* y, además, perfil del usuario, intenciones de búsqueda y ratio de *clicks* cada vez es más difícil que páginas o dominios completos que carecen de contenidos de calidad puedan quedar bien posicionados intentando los trucos habituales del *blackhat* o del *greyhat* SEO.

Esto nos lleva a otras dos conclusiones: aunque perdemos capacidad de predecir cómo van a quedar unas páginas determinadas en la SERP, sabemos que podemos seguir apostando por contenidos de calidad y por el resto de herramientas conceptuales, analíticas y técnicas para eliminar barreras a la indización, mejorar cada vez la estrategia de contenidos y dar cada vez mayor visibilidad a nuestros sitios.

La tercera conclusión es que, para los sitios web de medios de comunicación puede resultar muy fácil incorporar a su estrategia de contenidos lo que nos dicen los estudios sobre minería de consultas e intenciones de usuario.

Por ejemplo, para obtener tráfico para nuestro sitio podemos publicar guías y artículos tipo tutoriales y usar en los títulos los tipos de palabras o expresiones que hemos visto que indican búsquedas con posible motivación informacional.

Lo mismo podemos hacer con el resto de tipos de búsquedas. Podemos planificar una estrategia de creación y publicación de contenidos que tenga en cuenta, según el perfil de nuestro medio de comunicación, las palabras o expresiones que tienden a configurar búsquedas navegacionales o transaccionales y aplicarlos en los títulos de los contenidos y los metadatos.

En el caso concreto de búsquedas de motivación transaccional debemos ser conscientes que los enlaces patrocinados aparecerán con más facilidad y por tanto también adquieren relevancia, y que para obtener tráfico para nuestro sitio y, por lo tanto, conversiones, debemos posicionar nuestros contenidos no

solamente entre el Top10, sino a poder ser entre los tres primeros resultados orgánicos.

6. Bibliografía

- ▶ BARRY, C.; LARDNER, M. (2011): 'A Study of First Click Behaviour and User Interaction on the Google SERP. Recuperado de http://link.springer.com/chapter/10.1007/978-1-4419-9790-6_7
- ▶ BRODER, A. (2004): 'A taxonomy of web search'. Recuperado de <http://www.cis.upenn.edu/~nenkova/Courses/cis430/p3-broder.pdf>
- ▶ JANSEN, B. *et al.* (2008): 'Determining the informational, navigational, and transactional intent of Web queries', en *Information Processing & management*. Recuperado de http://faculty.ist.psu.edu/jjansen/academic/pubs/jansen_user_intent.pdf
- ▶ ROSE, D. E. y LEVINSSON, D. (2004): 'Understanding User Goals in Web Search'. Recuperado de <http://www.ambuehler.ethz.ch/CDstore/www2004/docs/1p13.pdf>
- ▶ WHITE, R. W. y DRUCKER, S. M. (2007): 'Investigating Behavioral Variability in Web Search'. Recuperado de <http://research.microsoft.com/pubs/68143/whitewww2007.pdf>

