

CASOS PRÁCTICOS DE INFORMATION SEEKING EN EL DISEÑO DE SISTEMAS DE INFORMACIÓN WEB

Jorge Serrano Cobos

Responsable del Departamento de Contenidos

MASmedios.com

jorgeserrano@gmail.com

Resumen: Se describen diferentes experiencias del estudio de la conducta de búsqueda de los usuarios, y su aplicación práctica en casos de éxito, obteniéndose aumentos significativos en términos de páginas vistas de secciones concretas y de portales en general, demostrando que una mejora en la organización jerárquica y visual de los links (browsing) permite una mejor recuperación de información y una menor utilización del buscador interno (searching) quedando posicionada ésta como segunda estrategia de búsqueda en el orden de preferencias de búsqueda de gran parte de los usuarios dentro de un espacio de información concreto.

Palabras Clave: Interaction Design, Information Seeking, User Search Behaviour, Search logs, Search Analytics, Information architecture. Diseño de Interacción, Conducta de Búsqueda de los usuarios, Arquitectura de Información

Abstract: Information seeking user behaviour on a web site can be analyzed through quantitative and qualitative studies using search logs. Different experiences involving search analytics are described. Also is studied its methodology, advantages for information seeking studies, key performance indicators, and its practical application to Information Architecture, improving Information Retrieval, internationalization of web sites, and as a help on seasonality future predictions using keywords search trends. Through several success stories, a better performance is observed, showing that a better hierarchical, alphabetical, or visual link organisation drives into a better information retrieval experience (browsing seeking), whereas internal search engine is less used, as a second option for information retrieval purposes (searching seeking) within several information spaces, integrating both user information search behaviour strategies

Keywords: Interaction Design, Information Seeking, User Search Behaviour, Search logs, Search Analytics, Information architecture

1 INTRODUCCIÓN

Una de las metodologías más interesante de estudio cuantitativo de las conductas de búsqueda de usuarios es a través del estudio de search logs, en los que se recoge todo lo que los usuarios buscan a través del buscador, en un sitio web. Es una disciplina muy aplicada al comercio electrónico, pero los resultados obtenidos de estos search logs no se comparten con la comunidad científica, puesto que se utiliza en entornos profesionales lógicamente muy cerrados, debido al carácter comercial y altamente sensible de ese tipo de análisis de información, tan útil para la cuenta de resultados de las empresas. Unida a la aplicación de tags a las páginas internas de un sitio web, podremos completar la información de la navegación de los usuarios y obtener más información tanto cualitativa como cuantitativa de la conducta de nuestros usuarios.

En esta ponencia se intentará desgranar algunas aplicaciones prácticas que este tipo de estudios tienen para la Arquitectura de Información de sitios web, y en general para la disciplina del Diseño Centrado en el Usuario (DCU), aquél que intenta diseñar sistemas de información teniendo al usuario como centro de sus objetivos.

2 INFORMATION SEEKING A TRAVES DE SEARCH LOGS

El estudio de la conducta de búsqueda de información (user search behaviour o information seeking) es el de aquella que se produce cuando un usuario interactúa con un sistema de recuperación de información específico para buscar información relevante (Wilson, T.D., 2000)

A su vez, Marchionini definió los dos procesos clave implicados en el proceso global del information seeking: la búsqueda directa y el browsing:

La búsqueda directa (searching) implica un esfuerzo para localizar un objeto conocido, una respuesta. El usuario tiene un concepto mental de aquello que busca, compuesto de características y propiedades intrínsecas que sirven de guía al usuario en su búsqueda.

El proceso de browsing, denominado por Marchionini discovery, (Marchionini, 1995), implica un esfuerzo para explorar un espacio de información susceptible de dar respuestas, en su búsqueda por objetos de conocimientos no conocidos ni especificados como en el caso de la búsqueda, en el que se examinan candidatos hasta encontrar los adecuados, y los aspectos relacionados con la percepción cuentan tanto como cuenta en el caso de la búsqueda directa la relevancia de las respuestas.

Marchionini aplicó su estudio a muchos tipos distintos de usuarios y demostró, coincidiendo con Koch (Koch, 2005), que los usuarios preferían en general las estrategias de browsing que la búsqueda directa (analytical strategies según Marchionini), independientemente de que variara enormemente sus habilidades. En los casos prácticos veremos cómo se han aplicado estos principios con un elevado éxito.

La metodología de estudio que nos ocupa, por contraposición a los tests de usuarios (que intentan conocer los pasos que los usuarios dan para encontrar información mediante el estudio de unos pocos usuarios representativos del universo de usuarios de un sistema de información dado), se centra en el archivo log del buscador, un archivo en el que se guardan todas las acciones llevadas a cabo por los usuarios en el servidor que ofrece los contenidos del sitio web.

Su gran virtud es que aporta más datos cuantitativos y es muy exhaustivo, porque el estudio abarca toda la población que utiliza el servicio mediante el archivo log del buscador, en el que se incluyen las palabras que los usuarios introducen en el sistema de recuperación de información de búsqueda directa. La mezcla de este archivo con el estudio de la navegación a través de tags o del log de visitas del servidor, nos permitirá conocer qué buscan los usuarios y qué ocurre antes y después de ese momento.

2.1 Datos cuantitativos

El search log informa de:

- lenguaje usado ("labelling" o etiquetado)

- quién y cuándo lo ha buscado (cuando el usuario se ha registrado y sus datos se tabulan junto a los términos que se buscan)
- del contenido encontrado,
- y del no encontrado.

Además permite:

- Aportar métricas exhaustivas
- Evaluar el lenguaje documental
- Saber cómo los usuarios definen un concepto ("modelo mental")
- Valorar el sistema de recuperación de información, el éxito (o no) del buscador (exhaustividad, pertinencia...)
- Valorar el grado de visibilidad del contenido

2.2 Beneficios para el Diseño Centrado en el Usuario

Los beneficios que puede aportar en la metodología de estudio del usuario son múltiples:

- Es un sistema no intrusivo para el usuario
- Sistema escalable
- Aporta datos cuantificables y sin tareas predefinidas
- Los datos son exhaustivos (de todos los usuarios del sitio)
- Los datos lo son del propio contexto del sitio
- Permite muchos análisis sincrónicos y diacrónicos

A partir de esos datos, podemos descubrir los conceptos relacionados que más se buscan, y ordenar los conceptos sabiendo cuáles son los que más se utilizan. De esta forma podemos normalizar el lenguaje documental acercándolo al lenguaje natural (Rosenfeld, 2002)

En cuanto a la arquitectura de información, permite conocer qué categorías serían apropiadas para organizar el contenido. Así, podríamos concertar que las categorías principales se forman con los conceptos más buscados, y las subcategorías con las palabras clave menos buscadas, o definir esas categorías principales mediante aquellas queries o búsquedas que contienen menos palabras, correspondientes a conceptos más generales, y subcategorías a través de las cadenas de búsqueda con más palabras.

Por tanto, si la findability es anterior a la usability ("no podemos usar lo que no podemos encontrar", Peter Morville) saber cómo buscan los usuarios (information seeking) será prioritario para entender qué instrumentos de recuperación de información vamos a preparar para nuestros futuros usuarios.

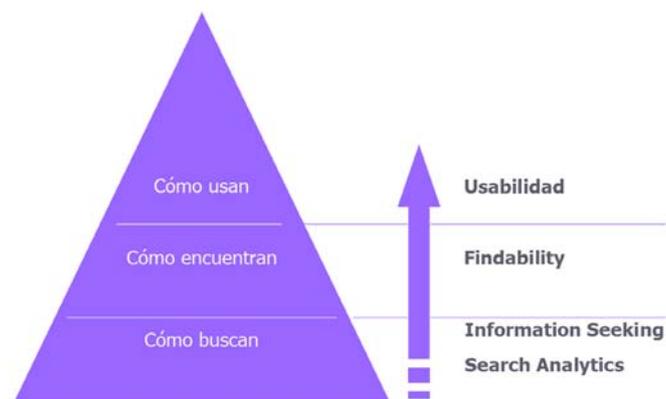


Fig.1 Proceso de Diseño Centrado en el Usuario basado en el estudio de conductas de búsqueda de los usuarios. (creación propia)

2.3 Limitaciones de los estudios de search logs

Sus limitaciones provienen de problemas técnicos sobre todo, y en algunos casos también metodológicos en cuanto que:

- Problemas en la reconstrucción de sesiones y caminos de navegación, debido a dificultades técnicas como páginas no cacheadas, uso de servidores proxy o firewalls, etc.
- Muchas IP no permiten reconocer el país del que procede un usuario
- No tenemos datos de la calidad de las conexiones de los usuarios, por lo que no podremos saber si el sitio funcionaba correctamente para ellos, o la navegación les resultaba lenta, lo que conlleva una variabilidad en el estudio que puede resultar influyente, sin conseguir una homogeneidad en la toma de datos a ese respecto.
- No podemos inferir si el usuario realmente encontró lo que necesitaba, sólo si encontró algo, lo visitó, incluso si lo consultó durante mucho tiempo y/o si lo volvió a visitar.
- No conocemos los intereses previos del usuario, ni los intereses reales que hay tras la expresión de su estado anómalo del conocimiento en lenguaje natural.
- ral, cada estudio basado en search logs se circunscribe a las características del sitio web que lo aloja, y la estructura de cada sitio web determina la conducta de búsqueda de los usuarios, (Kralisch y Berendt, 2004) por lo que las conclusiones de un estudio concreto pueden bien no ser plenamente extrapolables a otro sitio.

3 EXPERIENCIAS PREVIAS DE ESTUDIOS DE INFORMATION SEEKING EN SITIOS WEB.

Desde el primer momento de las investigaciones de este campo se han intentado clasificar los diferentes tipos de conducta de búsqueda, desde el punto de vista de la motivación que lleva a buscar. La clasificación mayoritaria de tipologías de conductas de búsqueda nos da conductas bien orientadas a un objetivo (goal oriented) bien búsqueda por exploración (exploratory search) Nielsen 1997, o Moe en 2003, realizaron estudios empíricos sobre estas tipologías de conductas de búsqueda.

Se han estudiado también otras variables que podrían afectar a la conducta de búsqueda, como por ejemplo hizo Shneiderman (1997) sobre el tipo de tarea a realizar durante la búsqueda. Scaife y Rogers (1996) utilizaron la presentación de la información, el interface, como variable dependiente en la conducta estudiada.

En la misma línea, Jones, Cunningham y McNab, de la Biblioteca Digital Nacional Neozelandesa, constataron que los usuarios raramente cambian las opciones marcadas por defecto, corroborando lo indicado por Nielsen al respecto, aunque este aspecto podría estar íntimamente relacionado con la disposición de las opciones de reconfiguración de la búsqueda.

En el estudio se cambiaron las opciones por defecto, sin que se produjera un aumento de reconfiguraciones, lo que induce a pensar que el usuario simplemente acepta la configuración por defecto como óptima.

En su experimento el tipo de operador booleano más usado fue el de intersección (AND, Y) seguido del de unión (OR, O), y después por el de negación (NOT, NO). Dependiendo del tipo de interfaz que se le ofrece al usuario, la proporción de elecciones de operador variaba, aunque respetando ese orden de prioridad de elección de operador, siendo menos usados los operadores booleanos cuando el algoritmo de ranking de respuestas ya los incorpora, lo que ocurre en el buscador Google, por ejemplo, el cual ya incorpora el operador NEAR por defecto.

En cuanto a la reformulación de la búsqueda, constataron que la mitad de consultas reformuladas se construía a partir del texto anterior, teniendo al menos un término en común, añadiendo o eliminando algún término. Este alto índice en la práctica de reformulación de búsquedas ya nos da una idea de la tendencia básica de uso de un buscador: consulta, revisión de las respuestas, reformulación de la búsqueda y nueva revisión, en iteraciones múltiples, hasta concluir el proceso, bien por abandono, bien por haber encontrado lo que se buscaba.

La aparición de resúmenes apropiados en las respuestas de las SERPs (Search Engine Result Pages, páginas de resultados de los buscadores) resulta de importancia significativa para discriminar entre respuestas que son útiles o no, confirmando la evolución en este aspecto que se produjo entre la disposición textual que se observaba en los buscadores tipo Altavista hasta 1998 y la que se ha ido imponiendo a partir de la aparición de Google, que siguió los consejos de Nielsen al respecto.

Jones y su equipo coinciden con otros autores y con mi propia experiencia en que el contenido disponible, algoritmo de recuperación e interfaz de consulta pueden modificar en ciertos aspectos la conducta de búsqueda de los usuarios, por lo que cada sitio web tendrá peculiaridades únicas, y las conclusiones de cada estudio pueden no ser siempre extrapolables a otros sitios web.

La variable más estudiada ha sido sin duda la de la capacidad cognitiva del usuario, en dos variantes: el grado de conocimiento de una temática por un lado, y el grado de experiencia en navegación a través de internet y de herramientas de búsqueda por otro. Por ejemplo, Navarro-Prieto 1999, Vakkari 2000.

En la misma línea, otros autores se han decidido por explorar los aspectos culturales de los usuarios y su incidencia en sus conductas de búsqueda de información, como Barber y Badre 1995, Sheppard y Sholtz 1999, Marcus 2000, Chau et al. 2000.

Borgman examinó las diferencias individuales que pueden modificar la conducta de los usuarios en material de recuperación de información, como características personales, aptitudes técnicas y orientación académica, y concluyó encontrando que estos factores estaban interrelacionados con su conducta. (Borgman, 1989)

Allen apuntó la idea de que el grado de conocimiento de una temática afectaba a la conducta de búsqueda de los usuarios (information seeking) en un catálogo de biblioteca online, así como a la efectividad de sus búsquedas en un sistema de información (search effectiveness o search performance). Los usuarios con mayor conocimiento usaban más expresiones de búsqueda hasta encontrar lo que buscaban en las SERPs. Allen señalaba que debería tenerse este factor en cuenta a la hora de diseñar sistemas de recuperación de información personalizados (Allen, 1991).

Kuhlthau comprobó en un estudio diacrónico (longitudinal studies) cómo a lo largo del tiempo, el juicio de relevancia de un usuario puede cambiar en el tiempo, reflejando su conocimiento personal de una temática y su comprensión del problema, puesto que lo que era relevante tiempo ha, puede ya no serlo después, y viceversa. En él, estudió un usuario durante 5 años para comprobar cómo cambiaba su conducta de búsqueda conforme mejoraba su conocimiento de los medios para realizar esas búsquedas.

Sources	Experience	
	Novice	Expert
Internal		
Newspapers and trade journals	X+	X
In-house specialists	X+	X
Reports of other analysts	X	0
External		
Annual reports	X	X+
Company briefings	X	0
Company visits	X	X+
Clients	0	X+
Institutional resources		
Libraries	X	X

* Code for participants rating of sources: 0 = little or no use; X = some use; X+ = frequent use.

Fig.2 Evolución temporal en el uso de fuentes de información. Kuhlthau, 1999.

La limitación de esta investigación es que sólo pudo ser realizada con un único sujeto investigado, por lo que debería ser estudiado más ampliamente, pero nos da una idea del cambio que se produce en el sujeto en variables como la del gráfico, en el que se observa cómo dependiendo de su experiencia en un campo temático, el usuario selecciona un tipo de fuentes u otras. Especialmente llamativo (y por otra parte, lógico) es el caso de los clientes, que en un estado inicial no se tienen en cuenta como fuente de información, y pasados 5 años, el usuario los considera una fuente de información vital en su proceso de búsqueda de información.

Marchionini et. al. en 1993 trabajó sobre la comparación en la conducta de búsqueda entre expertos en una determinada temática y expertos en técnicas de búsqueda. Su trabajo reveló que los expertos en una temática se fijaban más en el contenido de las respuestas y en cómo respondían a la pregunta o query formulada, mientras que los expertos en búsqueda de información se interesaban más por la formulación de la pregunta.

Anett Kralisch y Bettina Berendt, a partir de otros como Rice, McCreadie, y Chang, mostraron cómo la conducta de búsqueda de los usuarios podía estar determinada por circunstancias culturales, lo cual debía ser tenido en cuenta a la hora de diseñar las herramientas de recuperación de información internas de un sitio web. Propusieron así destacar unos servicios de búsqueda u otros, dependiendo de su nivel cultural.

Los usuarios con mayor nivel cultural demostraron tener mayor capacidad para absorber información contextual, y por lo tanto tendían más a utilizar los enlaces, organizados jerárquica o alfabéticamente, que aquellos usuarios con poca capacidad de absorción informativa, quienes rápidamente usaban el buscador.

Fidel y Efthimiadis observaron la conducta de búsqueda de información en internet de varios ingenieros en Boeing. Se constataron diversos patrones de búsqueda comunes:

todos los participantes en el experimento tomaron en consideración la relevancia y la confianza que ofrecía un sistema de información dado como los factores más importantes a la hora de decidir elegirlo para encontrar información sobre un tema concreto.

A la hora de reformular las búsquedas, los ingenieros suelen desarrollar búsquedas más específicas que generales.

Todos eligieron la facilidad de uso como el criterio más importante para decidirse por un método de búsqueda.

Cheuk y Dervin por su parte compararon 3 grupos distintos de usuarios: auditores, ingenieros y arquitectos. Confirmaron el modelo de Marchionini, pero a su vez, concordando con otros estudios anteriores, descubrieron diferencias de conducta dependiendo del tipo de usuario.

Sin embargo, Ellis y Haugan en un estudio con ingenieros y científicos noruegos habían identificado características similares: investigación, encadenamiento de búsquedas, monitorización de los resultados, browsing (hojeo), distinción de resultados relevantes, filtrado, extracción y finalización del proceso de búsqueda (Ellis y Haugan, 1997).

Vakkari y su equipo a su vez, (Vakkari, 2003) dirigieron un estudio en el que se investigó a 22 estudiantes de psicología mientras aumentaba su entendimiento de la materia y su experiencia en la interacción con buscadores. Su conclusión alude también al hecho de que el conocimiento de un dominio concreto del conocimiento tiene impacto sobre su conducta de búsqueda, aumentando el número de términos mientras aumentaba el grado de conocimiento de un tema dado (lo que coincide con lo indicado previamente por Allen) asumiendo que los usuarios tengan un conocimiento suficiente del sistema de recuperación usado.

Zhang, Anghelescu, y Yuan, trabajando en la línea marcada por Allen en 1991, constataron en 2005 que a medida que el nivel de conocimiento de un usuario sobre un determinado tema aumentaba, también aumentaba el número de búsquedas utilizadas para encontrar información que se ajustara a sus necesidades, así como el número de términos utilizados en las consultas o queries. Sin embargo, el nivel de efectividad (search effectiveness) en los resultados resultó el mismo para usuarios expertos en la temática y no expertos.

A tenor de los autores este hecho parece deberse al elevado número de entradas en el portal a través de enlaces internos y no de la home o portada, y a la disposición en pantalla de las herramientas de recuperación de información, que favorecía el browsing sobre la actividad de búsqueda mediante buscador (searching propiamente dicho), siendo en su mayoría secuencias seguidas de hasta 10 pasos en su actividad de browsing.

Otro aspecto del problema de la recuperación de información en el mundo web es el lingüístico. No es raro ya ver usuarios buscar información en diferentes idiomas, por lo que los SRI deben ser diseñados ya pensando en que vayan a ser usados por usuarios de distintas nacionalidades y cuyo lenguaje materno es distinto al común del SRI.

Kralisch y Berendt estudiaron en 2004 la conducta de búsqueda de usuarios de habla nativa y no nativa de un portal sobre medicina, para determinar el impacto que los problemas lingüísticos derivados del desconocimiento lingüístico de la lengua del portal podían producir en su conducta, al priorizar el uso de las diversas herramientas de recuperación de información habilitadas, tales como buscador interno, enlaces organizados jerárquicamente o enlaces organizados alfabéticamente.

Los resultados de su investigación apuntaron la conclusión de que el esfuerzo adicional requerido para usar las diferentes herramientas de recuperación de información tenía poco impacto en los usuarios no nativos, observándose similar conducta en unos y otros:

- En ambos casos, los enlaces organizados alfabéticamente fueron los más usados, y el buscador interno la opción menos usada.
- Asimismo, cuando se optaba por utilizar un sistema de recuperación de información concreto de los tres mencionados, frecuentemente se usaba de modo casi exclusivo, por lo que al usuario le costaba esfuerzo cambiar de sistema.
- Los contenidos organizados alfabéticamente se consideraron en ambos casos la alternativa por antonomasia cuando fallaba cualquiera de las otras opciones probadas en un primer momento de la navegación.

Sin embargo, la conducta de navegación de éstos sí era distinta de los usuarios nativos en los primeros momentos de la navegación. Los usuarios nativos de este portal médico utilizaban en mayor medida los contenidos organizados jerárquicamente. Los no nativos usaban en general más el buscador y los enlaces organizados alfabéticamente, lo que llevó a pensar a los autores que los nativos no eran grandes conocedores de la materia y los no nativos sí, idea que se corresponde con lo señalado por Allen en 1991, Vakkari et al., (2003), o Zhang, Anghelescu, y Yuan en 2005 en lo tocante a la relación entre los cambios en el grado de conocimiento de una temática y la conducta de búsqueda.

Angiolillo, ingeniero de usabilidad de la empresa Verizon, ya se quejaba de la poca atención que el estudio de los search logs suscitaba entre la comunidad científica. Utilizó tests de usuarios y focus group para contrastar las hipótesis a las que llegaba tras consultar el archivo de movimientos y palabras claves más buscadas de los usuarios en su buscador interno.

Verizon.com, sitio web empresarial dedicado a aplicaciones informáticas de pago de facturas y gestión de tarjetas de crédito, visitada por un millón de personas a la semana y con 30.000-40.000 búsquedas internas al día, mostraba en los archivos de su search log una interesante tendencia: la mayoría de los usuarios que buscaban, intentaban antes encontrar links que resultaran apropiados a su necesidad de información.

Si el usuario no los encontraba, entonces buscaba algo en el buscador interno. Además, aunque existiera un link adecuado, cuanto menor era la prominencia del link, es decir, cuanto más abajo en la página estuviera visible, más era buscado por los usuarios en la cajetilla del buscador. Este punto parece confirmado por las observaciones hechas con la herramienta de Eyetools.com, que nos muestra, web a web estudiada, que el ojo humano tiene en general tendencia a buscar con la mirada (seek) primero lo situado arriba a la izquierda de la pantalla, luego va bajando con tendencia a terminar en el punto de la pantalla situado abajo a la derecha de la pantalla.

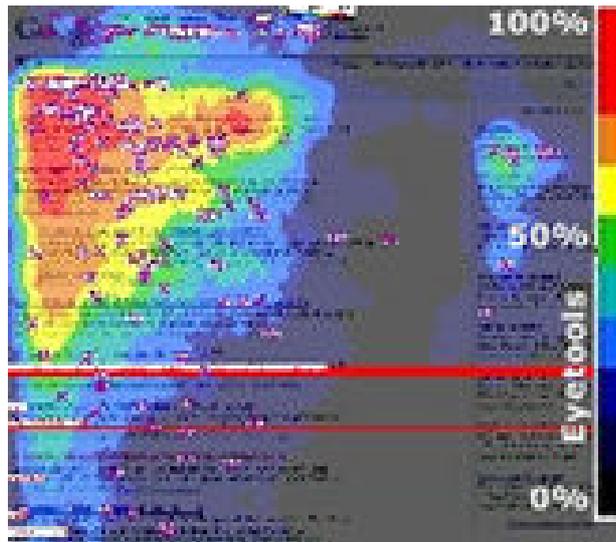


Fig. 4 Eyetools muestra con colores cálidos (rojo amarillo) zonas donde se fija más la vista. Eyetrack, 2005.

Por otro lado, Angiolillo constató que los usuarios noveles son los que más usaban el buscador interno, al contrario que los asiduos, que ya sabían adónde debían ir, y por lo tanto, centraban su atención en seguir los links adecuados, a los que ya se habían habituado.

Siguiendo la estela creada por Martin Belam en Currybetdotnet, Verizon utiliza la técnica de Best Bets (denominada aquí Best Matches) constatando otra aplicación práctica de la Ley de Zipf. Se tomaban las búsquedas más repetitivas y se preparaban un conjunto de respuestas a mano que según los documentalistas correspondieran mejor a la query o búsqueda formulada. Así, cuando un usuario buscaba algo, se le ofrecían dos juegos o subconjuntos de respuestas, las que ofrecía el algoritmo del buscador interno (basado en Verity) y las naturales. El subconjunto mejor aceptado por los usuarios fue el hecho a mano, los Best Bets.

Otros autores se han centrado en el estudio del uso de buscadores generales, como Google o Altavista. Así, autores como Clare Bradford e Ian Marshall denotaban que en las SERPs o páginas de respuestas de los buscadores, aparecía un elevado número de páginas duplicadas en sitios web diferentes o de contenidos idénticos en URLs distintas, alrededor de un 14% del total de páginas (Bradford y Marshall).

Estos duplicados ralentizaban la consulta de los listados de respuestas, al obligarlos a visitar sitios web que aparecían muy cercanas entre sí, como respuestas muy similares, descubriendo al visitarlas que eran el mismo contenido, lo que frustraba sus pesquisas, por lo que abogaban por el uso de URN (Uniform Resource Locator) para identificar unívocamente cada página.

Además, Clare Bradford e Ian Marshall estudiaron las estrategias de reformulación de búsquedas más comunes.

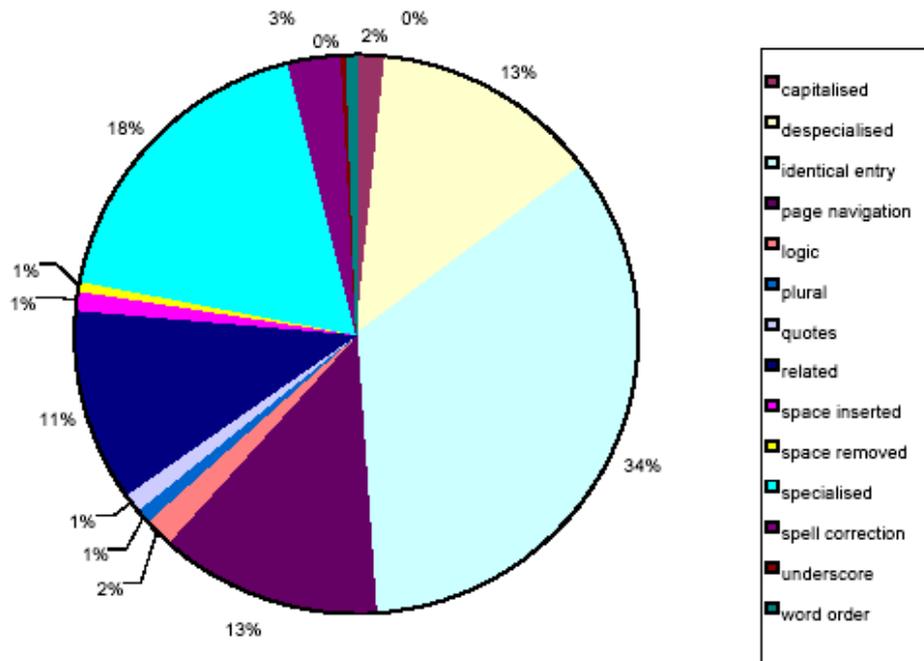


Fig. 5 Reformulación de búsquedas más comunes. Bradford y Marshall.

Descubrieron que, frecuentemente, el usuario abandonaba la búsqueda si la primera intentona no resulta satisfactoria.

Además, una búsqueda podía diferir enormemente de la siguiente, por lo que la personalización de agentes de búsqueda parece problemática.

Y por último, siguiendo a Zipf y a Pareto, comprobaron cómo el 50% del tráfico en buscadores se realizaba a través del 10% de búsquedas, y estas búsquedas son realizadas por un pequeño pero activo grupo de usuarios.

4 CASO PRÁCTICO DE APLICACIÓN A LA ARQUITECTURA DE INFORMACIÓN DE UN SITIO WEB

En el ejemplo que nos ocupa, se han requerido estudios de search logs para diseñar la arquitectura de información de un portal turístico, en dos momentos bien distintos: antes del desarrollo del mismo, y una vez desarrollado, en un proceso de rediseño. En ambos momentos podemos utilizar diferentes herramientas, tanto internas como externas, para conocer los hábitos de búsqueda de información a través de search logs y actuar en consecuencia.

En un primer momento, antes de haber desarrollado el portal, obviamente no tenemos información de nuestro search log interno. Para conocer las expresiones de búsqueda más comunes, podemos utilizar una herramienta como las que ofrece Google. En concreto se estudia el concepto “webcam”, para conocer qué expresión concreta es la más usada, y qué sinónimos o palabras relacionadas se buscan más.



Fig. 6. Google Adwords muestra por orden de más buscado a menos las expresiones de búsqueda sinónimas de una dada (en este caso “webcam”) y otras combinaciones

De esta forma podemos organizar el contenido, nombrar la sección (¿”webcam”, “webcams”, “web cam”...?), crear subsecciones, y preparar la información, metadatos, etc, para mejorar la findability interna de los contenidos y la visibilidad externa a través de su optimización para buscadores.

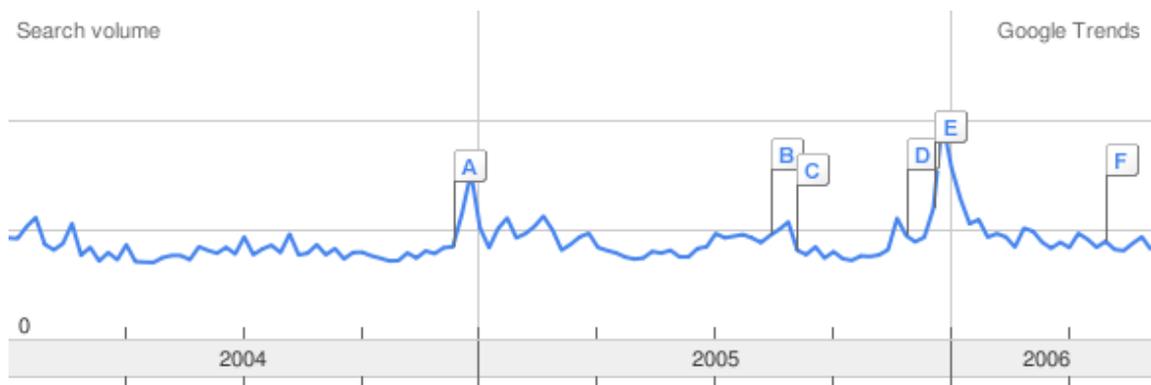


Fig. 7. Google Trends nos ofrece tendencias de búsqueda de un determinado concepto.

En el otro momento, durante el proceso de rediseño o de mejora del diseño, se pudo comprobar cómo el análisis de search logs nos puede ayudar a descubrir y solucionar problemas de navegación y estructura de información una vez el portal ha sido desarrollado.

Este portal cuenta con una lista de categorías ubicadas a la izquierda de la página, que se reproduce a lo largo de todo el sitio web. No todas las categorías aparecían en esa zona, sino sólo en la página inicial o home, y se sospechaba que eso producía problemas de recuperación de información y navegación entre los usuarios del portal.

Se consultó el search log del portal durante una serie de meses. Se encontró una constante en la lista de cadenas de búsqueda más frecuente de cada mes (top searched keywords). La constante fue que una de las principales búsquedas hacía referencia a una de las categorías generales, que sólo aparecía en la página principal, no en la lista de categorías que se utilizaba como recurso de recuperación de información mediante browsing jerárquico. En este caso el concepto era “webcam”, muy buscado constantemente. El problema era que un usuario que entrara directamente a una página interior mediante un enlace desde otra web o mediante un buscador, no sabía que existía la opción.

Recordando a Marchionini y los estudios de Eyetracking y combinando ambas tendencias, el intento de buscar una categoría en la izquierda y la reiterada mayor frecuencia de búsqueda, nos llevó a plantear la ubicación de la categoría, replicándola como otra más.

A partir de ese momento, se comprobaron de nuevo los search logs y los logs de visitas. Se constató un aumento de los niveles de visitas en la sección, aumentando el número de páginas vistas por usuario, y por otro lado una disminución del número de búsquedas para esa query concreta y otras asociadas semánticamente. Por otro lado, según los datos de OJD, observatorio independiente del tráfico de visitantes a sitios web a través de internet, observamos que se triplicaron el número de páginas vistas en la sección de webcams:

Así, la sección de webcams del portal en castellano, pasó de 192.267 páginas vistas, a 679.319 páginas vistas, lo que supuso un aumento de navegación en esa zona dentro del porcentaje de navegación total de la web, de un 7,09% al 21,13%.

Esta evolución, de cara a la usabilidad general del portal, confirma la validez de la hipótesis inicial: manteniendo las categorías generales a la izquierda durante todo el portal se aumenta el número de páginas vistas, y además, que aquellas opciones que se colocan más arriba y a la izquierda, se visualizan antes y tienen más posibilidades de ser clicadas.

5 CONCLUSIONES

- La Recuperación de Información es más que un algoritmo, engloba toda la experiencia del usuario.
- El análisis de palabras más buscadas debe ser una técnica fundamental del diseño centrado en el usuario.
- La unión del estudio de search logs y del log de visitas o sistemas de estudio de visitas mediante tags, no nos informan sólo del “searching”, sino también del “browsing”.
- A diferentes usuarios, diferentes soluciones de arquitectura de información (browsing y searching)
- Cuantas más posibilidades de recuperación de información, mejor.
- Las investigaciones futuras necesitarán implicar el cruce de técnicas y el uso de software de business intelligence, para mejorar la segmentación de usuarios en los futuros estudios de search logs.

REFERENCIAS BIBLIOGRÁFICAS

1. Allen, B.L. Cognitive research in information science: implications for design. En: Annual Review of Information Science and Technology, 1991b, 26, pp. 3-37.
2. Allen, B.L. Topic knowledge and online catalog search formulation. En: Library Quarterly, 1991a, 61(2), pp. 188-213.
3. Allerton Institute. How We Do User-Centered Design and Evaluation of Digital Libraries: A Methodology. Consultado en: 25-08-2006. <<http://edfu.lis.uiuc.edu/allerton/95/>>
4. Angiolillo, J. Search Log Analysis as a Usability Engineering Tool (word). Consultado en: 10-08-2006. <<http://home.earthlink.net/~searchworkshop/docs/chi-search-angiolillo3.doc>>
5. Borgman, C.L. All users of information retrieval systems are not created equal: an exploration into individual differences. En: Information Processing & Management, 1989, 25(3), pp. 237-251. Consultado en: 25-08-2006. <<http://informationr.net/ir/10-2/paper217.html#bor89>>
6. Bradford, C.; Marshall, I. Analysing Users WWW Search Behaviour (pdf) Consultado en: 15-07-2006. <<http://www.cs.kent.ac.uk/people/staff/iwm2/personal/ieewebcoll.pdf>>
7. Cheuk, W.; Dervin, B. A qualitative sense-making study of the information seeking situation: professionals in three workplace contexts. En: Electronic Journal of Communication, 1999, 9 (2-4). Consultado en: 15-07-2006. <<http://communication.sbs.ohio-state.edu/sense-making/art/artabscheukdervin99ejoc.html> >
8. Ellis, D. [et. al.] Information Searching. Part 5. User- intermediary Interaction. En: Journal of the American Society for Information. 2002. pp. 883-93.

9. Ellis, D. A behavioural approach to information retrieval design. En: *Journal of Documentation*, 1989, 46, pp. 318-338
10. Ellis, D.; Haugan, M. Modeling the information seeking patterns of engineers and research students in an industrial environment. En: *Journal of Documentation*, 1997, 53(4), pp. 384-403. Consultado en: 25-08-2006. <<http://informationr.net/ir/10-2/paper217.html#ell97>>
11. Fidel, R. Moves in online searching. En: *Online Review* 1985, 9 (1) pp. 61-74.
12. Fidel, R.; Efthimiadis, E. Web searching behavior of aerospace engineers. En: *Proceedings of the ACM International Conference on Research and Development in Information Retrieval*, 1999, 22, pp. 319-320. Consultado en: 4-09-2006. <<http://informationr.net/ir/10-2/paper217.html#fid99>>
13. Jones, S.; Cunningham, S.J.; McNab, R. Usage analysis of a Digital Library. Working Paper 98/13, Department of Computer Science, University of Waikato. Hamilton, New Zealand. 1998. (pdf) Consultado en: 10-08-2006. <http://www.cs.waikato.ac.nz/~stevej/Research/PAPERS/dl98logs.pdf>
14. Koch, T.; Ardö, A.; Golub, K. Log Analysis of User Behaviour in the Renardus Web Service. Consultado en: 10-08-2006. <<http://www.it.lth.se/knowlib/renardus-log-analysis.html>>
15. Kralisch, A.; Berendt, B. Cultural Determinants of Search Behaviour on Websites (pdf) En: *Proceedings of the IWIPS 2004 Conference on Culture, Trust, and Design Innovation*. Vancouver, Canada, 8 - 10 Julio, 2004. Consultado en: 25-08-2006. <http://www.wiwi.huberlin.de/~berendt/Papers/kralisch_berendt_CATAC2004_to_appear.PDF>
16. Kuhlthau, C.C. The Role of Experience in the Information Search Process of an Early Career Information Worker: Perceptions of Uncertainty, Complexity, Construction, and Sources. En: *Journal of the American Society for Information Science*. 50(5) pp. 399-412, 1999.
17. Mangano, V.; Beaulieu, M.; Robertson, S. Evaluation of Interfaces for IRS: Modelling End-User Searching Behaviour. En: *IRSG*. 1998.
18. Marchionini G. *Information Seeking in Electronic Environments*. Cambridge University Press, Cambridge, 1995.
19. Moe, W.; Buying, Searching, or Browsing: Differentiating between Online Shoppers Using In-Store Navigational Clickstream. En: *Journal of Consumer Psychology*, 2003, (13), pp. 29-39.
20. Navarro-Prieto, R.; Scaife, M.; Rogers, Y. Cognitive strategies in web searching. En: *Proceedings of the 5th Conference on Human Factors and the Web*. Julio, 1999.
21. Nielsen, J. *Designing web usability*. Indianapolis: New Riders, 1999.
22. Nielsen, J. Search and you may find. En: *Alertbox*, 15-07-1997. Consultado en: 25-06-2006. <<http://www.useit.com/alertbox/9707b.html>>
23. Rosenfeld, L. Search Log Analysis: The "Quiet" User Study. En: *User Studies and Information Architecture*. ASIST 2002 Annual Conference. 18-10-2002. (PPT) Consultado en: 02-05-2006. <<http://www.louisrosenfeld.com>>
24. Sheppard, C.; Scholtz, J. The Effects of Cultural Markers on Web Site Use. En: *Proceedings of the 5th Conference on Human Factors and the Web*, Julio, 1999.
25. Shneiderman, B. Designing information-abundant Web sites: issues and recommendations. En: Buckingham, S.; McKnight, C. (eds.) *Web Usability: special issue*, *International Journal of Human-Computer Studies*, 1997, 46.
26. Spink, A. Modelling Users' Successive Searches in Digital Environments, 1998. Consultado en: 25-06-2006. <<http://www.dlib.org/dlib/april98/04spink.html>>
27. Vakkari, P.; Pennanen, M.; Serola, S. Changes of search terms and tactics while writing a research proposal: a longitudinal case study. En: *Information Processing & Management*, 2003, 39(3), pp. 445-463. Consultado en: 25-08-2006. <<http://informationr.net/ir/10-2/paper217.html#vak03>>
28. Wildemuth, B.M. The effects of domain knowledge on search tactic formulation. En: *Journal of the American Society for Information Science & Technology*, 2004, 55(3), pp. 246-258. Consultado en: 4-09-2006. <http://informationr.net/ir/10-2/paper217.html#wil04>
29. Wilson, T.D. Human information behaviour. En: *Information Science*, 2000, 3(2), pp. 49-55. Consultado en: 15-07-2006. <<http://informationr.net/ir/10-2/paper217.html#wils00>>

30. Zhang, X.; Anghelescu, H.; Yuan, X. Domain knowledge, search behaviour, and search effectiveness of engineering and science students: an exploratory study. En: Information Research, Enero 2005, vol. 10 no. 2.