

Artículo

Análisis de estrategias de posicionamiento en relación con la relevancia documental

Por Jorge Morato, Sonia Sánchez y María Cruz Valiente

Resumen: El presente estudio trata de analizar distintas estrategias de posicionamiento en cuanto a su capacidad para obtener un nivel de exhaustividad óptimo. Las estrategias analizadas han consistido en la localización de los términos de búsqueda en el título, en la url, en los enlaces y en las etiquetas Meta y Head. El análisis se ha realizado bajo dos criterios. En primer lugar, se centra en páginas con escaso o nulo interés comercial. El motivo de eludir las páginas comerciales se debe a su estrecha vinculación a políticas de visibilidad y de optimización web. En segundo lugar, se valora el nivel de especificación de las consultas. El método de estudio ha consistido en realizar treinta con-



Jorge Morato es licenciado en CC. biológicas por la Universidad de Alcalá de Henares. A partir del año 1991 empezó a trabajar en distintas empresas relacionadas con la documentación. Obtuvo el título de doctor en documentación por la Universidad Carlos III en el año 1999. Desde ese mismo año ha impartido clases en las licenciaturas de documentación, LADE y en la ingeniería en informática en asignaturas relacionadas con la ingeniería de la información, recuperación de información e ingeniería del software.

Sonia Sánchez es licenciada en filología por la Universidad Complutense de Madrid y actualmente doctorando de documentación por la Universidad Carlos III. Desde el año 2002 es profesora ayudante en el departamento de informática de dicha Universidad con docencia en la licenciatura de documentación e informática sobre temas relacionados con modelado e ingeniería de la información. Actualmente se haya involucrada en distintos proyectos para la creación automática de ontologías.



sultas en los motores AskJeeves, Google, YahooSearch y MSN. Se han analizado los veinte primeros resultados de cada búsqueda determinándose su relevancia de forma manual. Los resultados determinan la utilidad de la estrategia del título; la necesidad de analizar los datos según el tipo de consulta; y la conveniencia de definir nuevos indicadores. Estos indicadores deberían considerar los documentos relevantes que carecen de la estrategia y los no relevantes que la contienen.

Palabras clave: Internet, Motores de búsqueda, Estrategias de posicionamiento, Medidas de precisión-recall, Relevancia documental, Algoritmo de posicionamiento,

Política de optimización, Hipótesis de la reina roja.



María Cruz Valiente es licenciada en ingeniería informática por la Universidad Carlos III y actualmente realizando el doctorado en informática. Es personal docente e investigador del departamento de informática desde el año 2003. Actualmente se encarga de la docencia relacionada con el análisis y modelado de dominios.

Title: Analysis of positioning strategies in accordance with relevance judgements

Abstract: This paper analyses the different positioning strategies according to their ability to obtain an optimum recall level. The analysed positioning strategies are the search for terms in the title, in the url, in the links, or in the Meta and Head tags. Our analysis is restricted to non-commercial pages –commercial pages are avoided because of their policies for achieving greater popularity by means of web optimisation-. The specificity level of the query is taken into account. We submitted thirty queries to four search engines (AskJeeves, Google, YahooSearch, and MSN); then we judged by hand the relevance of the first twenty hits of each query. Our analysis yields the following findings: the utility of the title strategy; the need for analysing the data by function of the type of query; and the suitability of designing new retrieval measurements. These measurements should take into consideration the relevant documents without the strategy, as well as those non-relevant documents with it.

Keywords: Internet, Search engines, Positioning strategies, Precision-recall measures, Relevance, Ranking algorithm, Search engine optimisation, Red queen hypothesis.

Morato, Jorge; Sánchez, Sonia; Valiente, María Cruz. “Análisis de las estrategias de posicionamiento en relación a la relevancia documental”. En: *El profesional de la información*, 2005, enero-febrero, v. 14, n. 1, pp. 21-29.

Artículo recibido el 25-10-04
Aceptación definitiva: 04-12-04

Introducción

La utilización y la popularidad de un motor de búsqueda se haya en función de la capacidad que tiene para seleccionar los documentos con mayor probabilidad de satisfacer las necesidades de información de un usuario y, con este objetivo, compiten por los mejores algoritmos de recuperación¹. Sin embargo, dado el volumen de información existente y los hábitos de los usuarios, que acostumbran a consultar los diez o veinte primeros registros de cada consulta, los algoritmos de recuperación no son suficientes². Por ello, los mo-

tores de recuperación web utilizan algoritmos de posicionamiento³. Basta realizar una búsqueda en cualquier motor para cerciorarse del interés que suscita la información sobre estos algoritmos.

El biólogo **Leigh van Valen** denominó *Efecto de la reina roja*⁴ a la coevolución entre especies, en la que el resultado de los cambios no supone una ventaja competitiva de una especie frente a otra. El término original fue tomado de *Alicia a través del espejo*. En el libro, la Reina comenta a Alicia “[...] en nuestro país necesitas correr todo lo que puedas para mantenerte en

Tipo de estrategia	Estrategias de posicionamiento
Localización y perfiles	- Ubicación geográfica del usuario, idioma, <i>cookies</i> , análisis de ficheros <i>.log</i> de servidores web, etc.
Fiabilidad	- Aparición en determinados dominios como <i>.gov</i> , <i>.edu</i> u <i>.org</i> . - Posición de la página html respecto a la raíz del dominio en la url.
Similitud de búsqueda	- Posición de los términos de búsqueda desde el inicio. - Proximidad de los términos y de frases de búsqueda en el texto. - Combinaciones basadas en la frecuencia del término (TF) y variaciones de la frecuencia documental inversa (IDF) ¹ y otros criterios asociados con TF como el tamaño de las páginas. - Presencia y prominencia de términos de búsqueda en el título, en enlaces, en el cuerpo del documento, etc. - Palabras clave al inicio y final de la página. - Conceptos asociados por clusters. - Url con el término de búsqueda.
Popularidad	- Hipertextual: enlaces entrantes a la página (p. e. <i>PageRank</i> de <i>Google</i> ⁸). - <i>Click-thru</i> . - Popularidad relativa dentro del sitio. - Estudio de visitas de la página (p.e. <i>reach</i> y <i>page views</i> de <i>Alexa</i>).
Novedad	- Preferencia a las nuevas incorporaciones y actualizaciones de páginas.
Código de la página y Formato	- Aparición y frecuencia en los metadatos y etiquetas <i>Alt</i> . - Formato <i>H1...H6</i> , negrita, tamaño de letra.
Usabilidad y accesibilidad	- Usabilidad en cuanto a factores como tiempo de carga, con pocas dependencias de software extras, etc.
Penalizaciones	- Utilización de técnicas de <i>cloaking</i> , texto oculto, repetición excesiva de términos, <i>cross-linking</i> .
Económicos	- <i>Pay-per-click</i> , <i>pay-for-inclusion</i> .

Tabla 1. Algunas variables utilizadas por los algoritmos de posicionamiento

el mismo sitio, para ir a algún sitio tendrías que correr por lo menos el doble de rápido[...]”. La competición entre algoritmos de posicionamiento y políticas de optimización de páginas provoca un efecto análogo al de la reina roja. Los algoritmos de posicionamiento necesitan reformularse continuamente para no ser manipulados por las técnicas de optimización de páginas⁵ que son aplicadas mayoritariamente en webs comerciales para aumentar así su visibilidad^{6,7}. Al margen de dicho enfrentamiento se encuentra un número importante de recursos que no tienen finalidad comercial, pero que se ven afectadas por dicho efecto al no aplicar políticas de optimización. Además, esta evolución causa a menudo que estrategias de posicionamiento muy eficaces no sean justamente valoradas por los algoritmos de posicionamiento. Previsiblemente, dicha coevolución podría perjudicar a la inmensa mayoría de las páginas que no están optimizadas.

En el presente documento se pretende diseñar un método que refleje el comportamiento de distintas estrategias de posicionamiento según el tipo de consulta y motor empleado. El objetivo del estudio no consiste en saber cuál es la estrategia que mejora en mayor medida la posición en un determinado motor, sino saber cuál de ellas nos permite recuperar una mayor proporción de documentos relevantes. Dicho de otro modo, algunas de las estrategias seleccionadas podrían no ser utilizadas por los algoritmos de los motores, ya que lo que se intenta analizar es si dicha estrategia resulta eficaz en la recuperación de documentos relevantes. Por otro lado, se discutirán posibles indicios del *Efecto de la reina roja* en las estrategias de posicionamiento.

«La competición entre algoritmos de posicionamiento y políticas de optimización de páginas provoca un efecto análogo al de la Reina Roja»

Se denominarán, de esta forma, políticas o técnicas de optimización a las estrategias y herramientas utilizadas para mejorar la posición o ranking de determinada página por parte de un motor de búsqueda concreto. Por otro lado, el término algoritmo de posicionamiento se reserva para el conjunto de variables que el motor de recuperación emplea para ordenar los resultados. Por último, se entiende estrategia o técnica de posicionamiento a todo factor que pueda influir en la ordenación de los resultados de búsqueda.

A continuación se ofrece una breve panorámica sobre las estrategias de posicionamiento y optimización básicas, no se trata de que sea exhaustiva, sino ilustrativa del estado actual. Posteriormente, se describe la metodología empleada y los indicadores diseñados

para realizar el estudio. En el siguiente apartado se muestran los resultados y observaciones obtenidas para, por último, pasar a discutirlos.

Estrategias de posicionamiento y optimización

Los algoritmos de posicionamiento utilizan con frecuencia un enfoque multivariante. En este caso, la combinación de distintos criterios se utiliza para calcular el peso por el que serán ordenadas las páginas frente al resto de resultados. En la tabla 1 se muestran algunas de las variables más utilizadas.

Existen diversas técnicas de optimización de páginas⁹. Destacan por su popularidad los estudios de los ficheros *.log* en los de buscadores, los programas de intercambio de enlaces y las granjas de enlaces (*link farms*, son acuerdos entre varios sitios web para incluir la misma lista de enlaces en cada uno de ellos), las herramientas para seleccionar páginas web¹⁰; las páginas para calcular la densidad de una sentencia de búsqueda¹¹; los simuladores de motores¹²; las aplicaciones para medir la popularidad de los enlaces entrantes y el tráfico a una web¹³.

Metodología

Con este estudio se pretende analizar varios aspectos diferentes:

—Concretar el grado en que las técnicas de posicionamiento sirven para localizar documentos relevantes y qué posibles indicadores podrían ayudar a su estudio.

—Determinar si la efectividad de estas estrategias depende del motor por el cuál se ha localizado.

—Analizar si el resultado está afectado por el tipo de consulta realizada.

Para el primer objetivo el punto de partida ha sido una serie de consultas realizadas en diferentes buscadores, llevadas a cabo todas ellas por estudiantes del programa de doctorado en informática. Se evitaron aquellas que pudieran recuperar un gran porcentaje de páginas comerciales y ellos mismos valoraron manualmente la relevancia de los veinte primeros resultados.

Las consultas se han clasificado en dos tipos:

—Suficientemente especificadas (p.e. “cuidados de la planta Gerbera”), o por el contrario

—insuficientemente especificadas (p.e. “Torrenueva”).

Se ha diferenciado entre este tipo de consultas para el análisis de los resultados, pero no se tiene como objetivo valorar su calidad.

Todas las variables se han almacenado de modo binario:

—Valor 1 para la presencia del término en la sección analizada.

—Valor 0 en su ausencia.

Con el fin de saber la eficacia de algunas de las estrategias reseñadas anteriormente (tabla 1) se ha estudiado la relevancia de los documentos dependiendo de si el término aparece, dentro de la página html, en:

—El título (*Tit*).

—En los metadatos (*Meta*)⁴.

—En los enlaces (*Link*).

—En la url (url).

—En las cabeceras (*H1,...H6,...Hn*).

La elección de estos indicadores, y no otros, está motivada por dos razones:

—Su extendido uso como elemento de posicionamiento en la mayoría de los motores de recuperación (con alguna excepción como *Google* con las etiquetas *Meta*⁴). Téngase en cuenta que en el presente análisis no se pretende saber si un motor utiliza determinada estrategia, sino si entre los primeros resultados devueltos se encuentran documentos relevantes que presentan la estrategia. En cualquier caso en toda la literatura sobre posicionamiento se recomienda poner etiquetas *Meta*, aunque el motor no las utilice. El motivo es, entre otros, poder ver una descripción del recurso (mediante la etiqueta *Meta description*) o poder adaptarse a una posible evolución de los motores.

—La facilidad para estimar sus valores informáticamente.

De esta forma, se ha procedido a evaluar el grado de eficacia de las estrategias en: *Yahoo! Search*, *Google*, *AskJeeves* y *MSN Search*. Esta elección se debe a que, a mediados del año 2004, todos ellos estaban clasificados entre los cinco más utilizados^{15,16}. Por el contrario, la ausencia de otros también conocidos se debe o bien a su baja utilización o por estar integrados en los ya señalados¹⁷.

«El enfrentamiento entre políticas de optimización de las páginas comerciales y la mejora de algoritmos de posicionamiento son perjudiciales para las páginas no comerciales»

Se han realizado treinta consultas y se han analizado los veinte primeros resultados retornados por cada motor ante cada una, estudiando el grado de utilización

por el motor de cada estrategia y la presencia de ésta en los documentos relevantes recuperados. Esta técnica es una adaptación de la propuesta por **Chignell** [et al.] para evaluar la recuperación en la web¹⁸.

Indicadores utilizados para el análisis de las estrategias

1. Exhaustividad de la estrategia

Es el número de documentos relevantes que utilizan la estrategia dividido por el número de documentos relevantes total:

$$Exhaustividad_{est} = |Relevantes_{est}| / |TotalRelevantes|$$

La ecuación, análoga al cálculo de la exhaustividad, indica qué proporción de los documentos relevantes tiene presente la estrategia. Su variación oscila entre cero (ninguno relevante) y uno (todos los relevantes tienen la estrategia). Esta ecuación ha sido empleada en los gráficos de precisión-exhaustividad (PR) que se muestran en los próximos apartados.

2. Precisión de la estrategia

Se calcula dividiendo el total de documentos relevantes con la estrategia entre el total de documentos analizados (veinte en nuestro caso).

$$Precisión_{est} = |Relevantes_{est}| / |TotalRecuperados|$$

Como en el caso precedente esta fórmula es equivalente al cálculo de la precisión pero limitado a la presencia de la estrategia. Indica entre los documentos recuperados cuántos han sido relevantes con la estrategia presente. Sus límites también estarían entre cero y uno.

3. NR

Es el número de documentos no relevantes que utilizan la estrategia dividido por el número de documentos no relevantes.

$$NR = |\neg Relevantes_{est}| / |Total\neg Relevantes|$$

Como se puede observar en los dos indicadores anteriores, para determinar la bondad de una estrategia no se tienen en cuenta determinados factores. Por ejemplo, los documentos no relevantes que contienen la estrategia o los que sí lo son pero no la contienen. NR y la siguiente ecuación tratan de resolver este problema. NR tiene una variación entre 0 y 1, y está relacionada con medidas como el *fallout*¹⁹ o el *IRC* de **Skrop**²⁰. Como en las medidas reseñadas, esta fórmula ofrece una medida de la recuperación aún en el caso de que se careciera de documentos relevantes. Este factor determina la probabilidad de que un resultado no relevante contenga la estrategia. Un valor bajo resulta positivo en la valoración de la estrategia.

	Exhaustividad de la estrategia	Precisión de la estrategia	NR	RT
Tit	0.7	0.3	0.5	0.3
Url	0.3	0.2	0.4	-0.3
Hn	0.2	0.1	0.0	-0.6
Link	0.4	0.2	0.4	-0.3
Meta	0.4	0.2	0.3	-0.3

Tabla 2. Indicadores de eficacia de la estrategia

4. RT

Es el resultado de calcular la diferencia entre los documentos relevantes con la estrategia menos los relevantes sin la estrategia entre el total de documentos relevantes.

$$NR = (|Relevantes_{est}| - |Relevantes_{\neg est}|) / |TotalRelevantes|$$

RT considera un factor no tenido en cuenta en las anteriores funciones y calcula los documentos relevantes que carecen de la estrategia. Su variación es entre -1 y 1. Un dato negativo parece indicar que la estrategia es perjudicial, y un valor más próximo a uno indica un funcionamiento óptimo.

Resultados

A continuación se analizan los datos obtenidos. Para ver cómo afecta el ranking a las estrategias se han realizado gráficos de precisión-exhaustividad (PR)¹. Si bien en algunos casos, para facilitar las comparativas, se ha extrapolado el valor al tanto por uno en vez de a la curva de precisión. Esto es, la cantidad de relevantes recuperados con y sin la estrategia siempre suman uno.

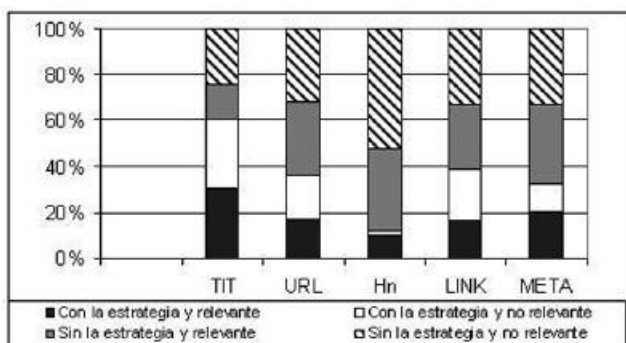


Figura 1. Relevancia porcentual según el conjunto de estrategias

Estudio de los factores que afectan a la valoración de una estrategia como relevante

En el gráfico de la figura 1 se muestra el comportamiento de las distintas estrategias según dos parámetros:

- Documentos relevantes o irrelevantes.
- Documentos con la estrategia presente o ausente.

Si se observa la figura 1 se puede deducir que:

—El título es la estrategia más prometedora en cuanto a la proporción de documentos relevantes que son recuperados. Sin embargo, es conveniente percatarse de la proporción de documentos con dicha estrategia y no relevantes.

—Las peores puntuaciones las obtiene *Hn* aunque sin embargo es, junto con las etiquetas *Meta*, el que presenta menor número de documentos no relevantes con la estrategia, lo cual apoya la tesis de que se trata de buenas estrategias.

—Salvo el título, todos recuperan una gran cantidad de documentos relevantes sin la estrategia.

—Por último, la ausencia de estrategia y de relevancia alcanza grandes proporciones en *Hn*. Este dato puede considerarse, en cierto modo, un aspecto positivo para valorar la estrategia.

En la tabla 2 se tienen en cuenta los factores definidos en el apartado de métodos.

En lo relativo al título, la exhaustividad corrobora los resultados de la figura 1. Es decir, el título obtiene gran número de documentos relevantes con la estrategia. El segundo mejor resultado corresponde al texto de los enlaces.

—La precisión obtiene un valor más alto en título.

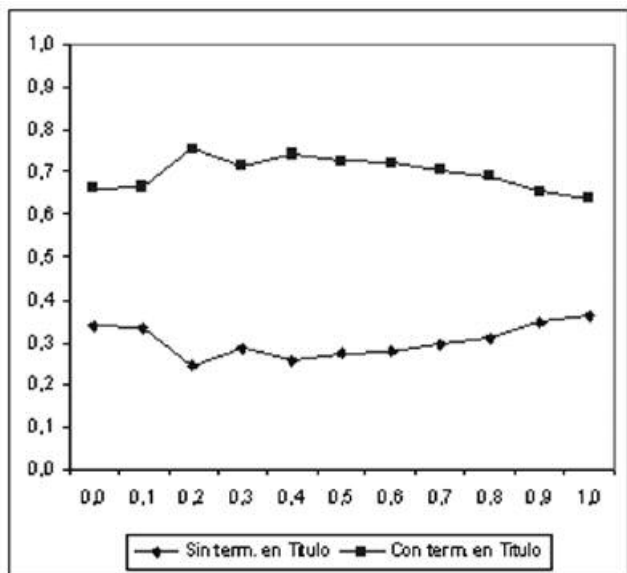


Figura 2 A

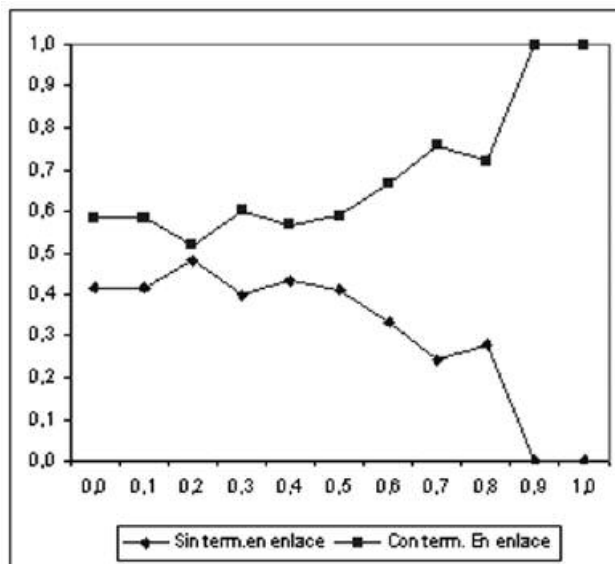


Figura 3 A

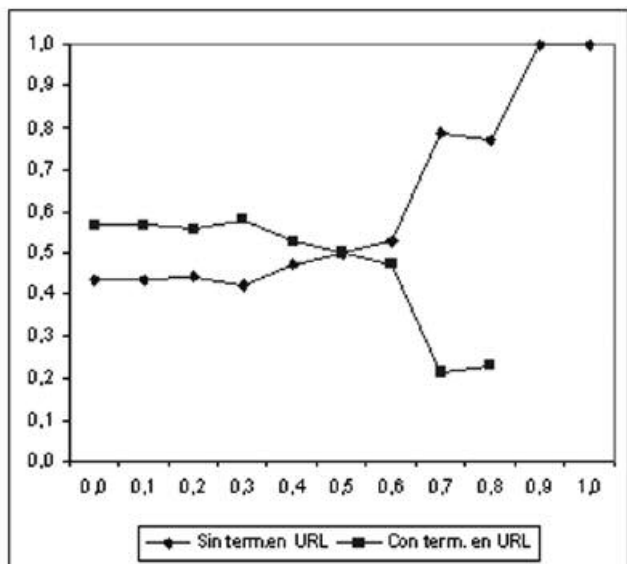


Figura 2 B

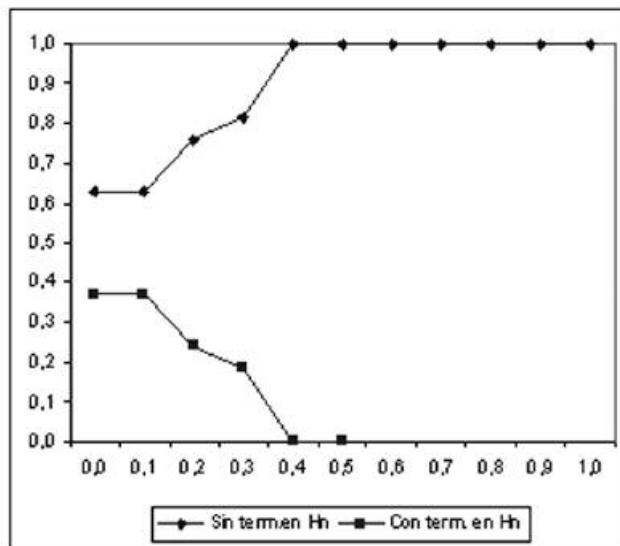


Figura 3 B

—Resulta significativo que la NR tenga un valor alto en el título. Este dato realmente podría estar señalando que el hecho de que los términos de la consulta se encuentren en él es un fenómeno muy probable. También resulta significativo el comportamiento óptimo de *Hn*.

—Por último, hay que destacar que RT tiene hasta cuatro valores negativos para las estrategias estudiadas, lo que indica que existen más documentos relevantes sin la estrategia que con ella.

Técnicas de posicionamiento y medida de la relevancia

Para ampliar el estudio iniciado en el apartado previo, se ha creído conveniente concretar los resultados para los documentos relevantes según contengan o no la estrategia a los once niveles de exhaustividad. Los gráficos de precisión-exhaustividad para las estrate-

gias estudiadas se muestran en las figuras 2 (A, B), 3 (A, B) y 4. Los gráficos han sido modificados ajustándolos a una precisión del tanto por uno para facilitar la comparación.

—Figura 2. A: gráfico PR con la variación de los documentos relevantes que contienen el término en el título y los relevantes que no tienen la estrategia. En el gráfico B se representan dichos resultados para la estrategia de url.

—Figura 3. A: gráfico PR mostrando los documentos relevantes con los términos de búsqueda en los enlaces y los que no los tienen. El gráfico B muestra los datos correspondientes a la estrategia de *Hn*.

—Figura 4: gráfico PR mostrando los relevantes con los términos de búsqueda bajo etiquetas *Meta* y relevantes sin la estrategia.

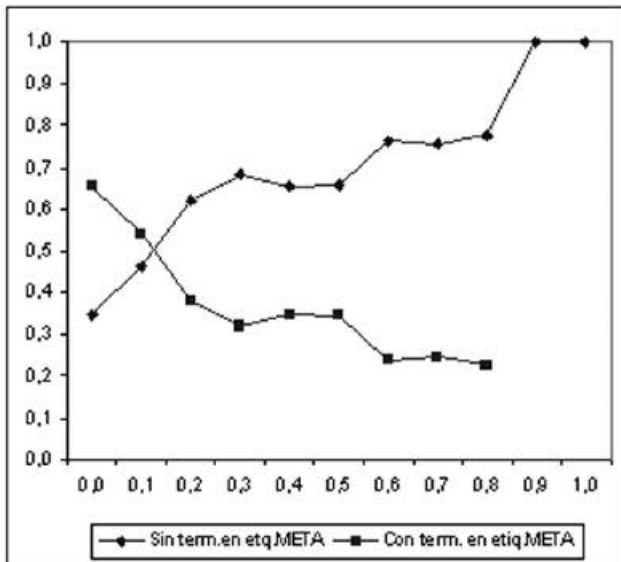


Figura 4

De la comparativa se han obtenido las siguientes observaciones:

—El título parece tener un comportamiento estable en los once niveles, presentando mejor resultado de precisión los documentos con la estrategia.

—En el resto de gráficos se observa que en los niveles de exhaustividad hasta el 20% tiene un mejor comportamiento la presencia de la estrategia (a excepción de *Hn*). Posteriormente, los resultados parecen tener una mayor proporción de documentos sin la estrategia que con ella. En las estrategias de enlace y url esta tendencia, aunque presente, está menos acusada.

Técnicas de posicionamiento y medida de la relevancia según los buscadores

Otra cuestión, que pretende explorar el presente estudio es en qué medida los distintos buscadores producen observaciones diferentes a las apuntadas en el anterior apartado. Este punto sólo incluye a dos de los cuatro propuestos: *Google* y *MSN*. El motivo de esta li-

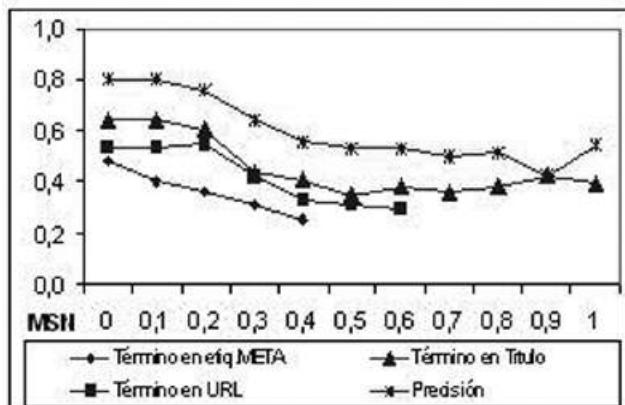


Figura 6. Gráfico PR para el motor MSN ante las distintas estrategias

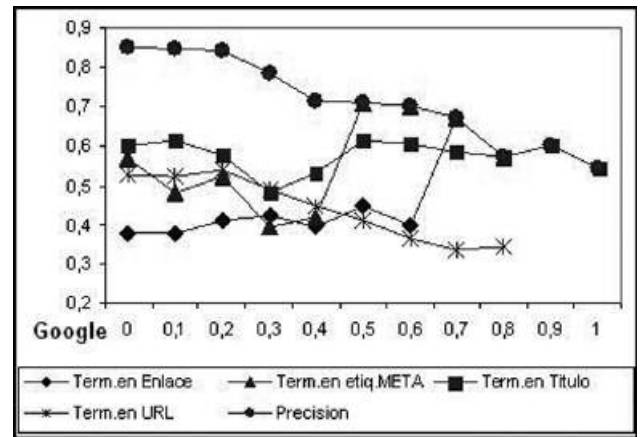


Figura 5. Gráfico PR para el motor Google ante las distintas estrategias

mitación se ha debido a la comprobación visual de que todos los motores presentaban una dinámica similar.

Una precisión necesaria en cualquier estudio que compare la exhaustividad entre buscadores es la ausencia de una única base de datos. Idealmente, cualquier comparación de este tipo debería partir de una base de datos única contra la que ejecutar los algoritmos de recuperación y posicionamiento.

Las figuras 5 y 6 muestran, respectivamente, los gráficos de PR para los buscadores *Google* y *MSN*. En este caso no se han extrapolado los resultados al tanto por uno sino que se conserva la gráfica de PR normal.

En la figura 5 se observa cómo algunas de las probables estrategias empleadas por *Google* para optimización parten de bajos puestos de exhaustividad (p.e. en los casos de la url y los enlaces). Este hecho se puede explicar en gran medida por las técnicas de optimización web que permiten posicionar bien documentos que probablemente no puedan resolver satisfactoriamente las necesidades de información del usuario. Las etiquetas *Meta* (según la literatura no consideradas por *Google*¹⁴) parecen tener un elevado peso en los primeros niveles para después desaparecer. De nuevo se observa una gráfica con tendencia ascendente (porcentualmente) en el título.

En *MSN* (figura 6) se observa una dinámica similar a la de *Google*, con un buen comportamiento para la estrategia de título. Significativamente, las etiquetas *Meta* parecen comportarse de un modo similar al percibido en *Google*.

Técnicas de posicionamiento y medida de la relevancia según el tipo de consulta

Por último, se ha intentado explorar en qué medida los resultados podrían verse afectados por la tipología de la consulta. Ya se ha comentado que han sido definidas como “correctamente especificadas” si no

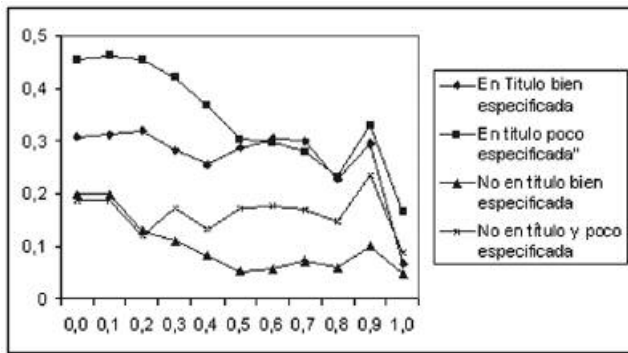


Figura 7. Gráfico PR para estrategia de título en consultas suficiente e insuficientemente especificadas

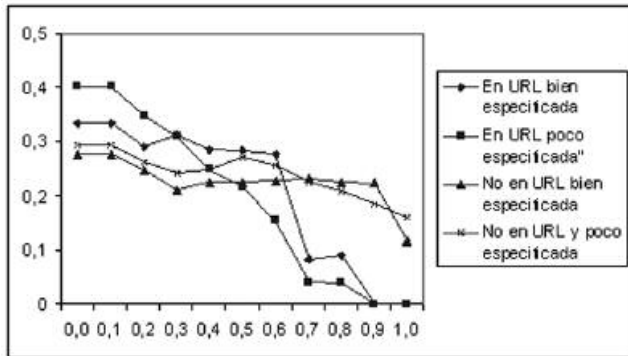


Figura 8. Gráfico PR para estrategia url en consultas suficiente e insuficientemente especificadas

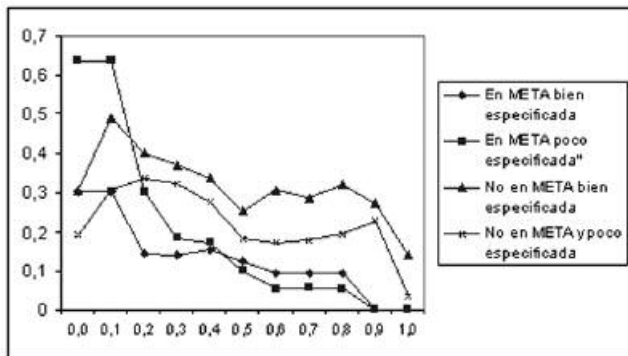


Figura 9. Gráfico PR para estrategia de etiquetas Meta en consultas suficiente e insuficientemente especificadas

daban lugar a apreciaciones subjetivas en cuanto a la valoración de un documento como relevante; o bien “consultas insuficientemente especificadas” cuando, por su formulación, contuvieran un alto grado de subjetividad para determinar su relevancia.

Se muestran así los gráficos PR con la precisión modificada a tantos por uno para facilitar la comparación (figuras 7 a 9).

De la comparativa se extraen las siguientes observaciones:

—La escasa diferencia entre los tipos de consulta en el caso de la url.

—El mejor comportamiento de las consultas mal especificadas en casi todos los casos.

Conclusiones

Del estudio realizado se han podido extraer las siguientes conclusiones:

—La hipótesis de partida de este estudio formula que el *Efecto de la reina roja* entre políticas de optimización y evolución de los algoritmos de posicionamiento provoca una situación de equilibrio entre las páginas que las aplican para su recuperación, pero a su vez se producen unas circunstancias desfavorables para aquellas que no las implementan. Este hecho se puede corroborar con el comportamiento de la estrategia de las etiquetas *Meta* y la url.

—Parece insuficiente medir la eficacia de una estrategia limitándose a la precisión y a la exhaustividad. En este documento se han propuesto nuevos indicadores que contemplen los documentos relevantes sin la estrategia y los no relevantes con la estrategia.

—El título parece ser la estrategia más robusta bajo todos los criterios estudiados.

—Al analizar los gráficos de PR, se observa que los documentos relevantes con la estrategia son más frecuentes hasta recuperar el primer 20% de documentos relevantes. A partir de ese punto, la tendencia tiende a invertirse. El motivo puede ser el cálculo multivariante en la asignación de pesos, aplicados por la mayoría de los algoritmos de posicionamiento. En un principio, los pesos encuentran los documentos más claramente relevantes, a partir de determinado punto no existe una tendencia tan clara, por lo que la estrategia no es tan determinante. Por otra parte, también podría deberse a que ese 20% corresponda a los documentos optimizados; de ser cierta esta hipótesis sería un nuevo indicio del *Efecto de la reina roja*.

«El objetivo del presente trabajo no consiste en saber cuál es la estrategia que mejora nuestra posición, sino la que permite recuperar una mayor proporción de documentos relevantes»

—El comportamiento de los buscadores, en lo relativo a la relevancia, parece estar afectado por las políticas de optimización en los primeros niveles de exhaustividad.

—La explicación al hecho de que las preguntas escasamente especificadas se comporten mejor en cuanto a la exhaustividad, podría ser:

a. Naturaleza de las consultas: frecuentemente, las poco especificadas contienen una o dos palabras, mientras que otra bien especificada consiste en una ca-

dena de términos cuya unión concreta el sentido. En el primer caso cualquier aparición del término en una página la puede convertir en un posible resultado válido; en el segundo, factores como proximidad de términos de búsqueda entran en juego.

b. El rigor en la valoración de la pertinencia: ante una consulta bien especificada el evaluador suele ser más estricto que ante una que lo está mal, ya que su necesidad de información se encuentra más definida.

Determinar cuál de estas dos razones es la que tiene un mayor peso en la explicación de las observaciones queda pendiente para futuros estudios.

Referencias

1. **Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier.** *Modern information retrieval*. New York: ACM Press; Harlow; Madrid [etc.]: Addison-Wesley, 1999. Isbn 020139829X.
2. **Sullivan, D.** *Search engine watch*. Consultado en: 01-10-04. Actualizada en: 02-09-03.
<http://searchenginewatch.com/reports/article.php/2156481>
3. **Marckini, Frederick.** *Search engine positioning*. Plano, Texas: Wordware Publishing Inc., 2001. Isbn 155622804x.
4. **Valen, L. Van.** "A new evolutionary law". En: *Evolutionary theory*, 1973, v. 1, pp. 1-30.
5. **Nobles, R.** *The future of search engine optimizing*. Consultada en: 01-10-04. Actualizada: 07-06-01.
http://www.searchengineguide.com/aws/2001/0607_aws1.html
6. **Molenaar, M.** *Search engine secrets. Yuma Business Direct, 2001*. Consultado en: 01-10-04.
http://www.mgmdesign.com/showarticle.html?article_id=53
7. **Bhargava, Hemant K.; Feng, J.** «Paid placement strategies for internet search engines». En: *Proceedings of the eleventh international conference on world wide web*, 2002, pp. 117-123. Isbn 1-58113-449-5.
8. **Page, L.; Brin, S.; Motwani, R.; Winograd, T.** "The PageRank citation ranking: bringing order to the web", technical report, Computer Science Dept., Stanford Univ., 1998.
9. **Azlor, S.** *Posicionamiento en buscadores: guía básica*. Consultado en: 01-10-04. Actualizado: 2003.
<http://www.guia-buscaadores.com/posicionamiento/>
10. *Espotting media-Delivering business to businesses*. Consultado en: 01-10-04.
<http://www.espotting.com/popups/keywordgenbox.asp>
11. *Keyword counter-Keyword frequency analyzer*. Consultado en: 01-10-04.
<http://www.keywordcount.com/>
12. *Search engine spider simulator*. Consultado en: 01-10-04.
<http://www.webconfs.com/search-engine-spider-simulator.php>
13. *Alexa web search*. Consultado en: 01-10-04.
<http://www.alexa.com>
14. **Sullivan, D.** *Death of a meta tag*. Consultado en: 01-10-04. Actualizada en: 01-10-02.
<http://searchenginewatch.com/sereport/article.php/2165061>
15. **Sullivan, D.** *Hitwise search engine ratings*. Consultado en: 01-12-04. Actualizada en: 17-05-04.
<http://searchenginewatch.com/reports/article.php/3099931>
16. **Sullivan, D.** *Nielsen netratings search engine ratings*. Consultado en: 01-12-04. Actualizada en: 14-07-04.
<http://searchenginewatch.com/reports/article.php/2156451>
17. **Notess, Greg R.** "The new Yahoo! Search". En: Online, 2004, July/August, v. 28, n. 4. Consultado en: 01-12-04.
<http://www.infoday.com/online/jul04/OnTheNet.shtml>
18. **Chignell, M. H.; Gwizdka, J.; Bodner, R. C.** "Discriminating meta-search: a framework for evaluation". En: *Information processing and management*, 1999, v. 35, n. 3, pp. 337-362.
19. **Rijsbergen, C. J. van.** *Information retrieval*. Butterworth, London, 1979. Isbn: 0408709510.
20. **Skrop, Adrienn.** *Irrelevance rate function: a complementary measure of web search engine effectiveness*, 2002. Consultado en: 01-10-04.
<http://citeseer.ist.psu.edu/skrop02irrelevance.html>

Jorge Morato, Sonia Sánchez, María Cruz Valiente, Departamento de Informática, Universidad Carlos III de Madrid.
jorge@ie.inf.uc3m.es
ssanche@ie.inf.uc3m.es
mvalien@ie.inf.uc3m.es

Versión online de EPI

Existe una versión electrónica de *El profesional de la información*, de uso gratuito para la mayoría de los suscriptores (empresas, organismos, instituciones), que pueden acceder a través de internet a los textos completos y materiales gráficos publicados en la revista.

Más información en:

<http://www.elprofesionalde lainformacion.com/contenidos.html>