

Aproximación a la evaluación cuantitativa de los sistemas de recuperación de información de la prensa en Internet: Exhaustividad y precisión

Juan Carlos García Gómez

Universidad de Murcia

José Luis González Olivares

Biblioteca Pública Municipal de Jumilla

0.1. Resumen

Las ediciones de los medios de comunicación escrita en Internet son cada vez más diferentes de las tradicionales, gracias a la incorporación de nuevos servicios que aprovechan las posibilidades que la red les ofrece: inmediatez, alcance mundial, hipertextualidad, multimedia, integración de bases de datos, flexibilidad de formatos, etc. Entre estos servicios destacan la personalización de la página principal de acceso, el envío personalizable de noticias por correo electrónico y los sistemas de recuperación de noticias, cuyas características se evalúan tanto cuantitativamente mediante el modelo de evaluación de recuperación de la información, como cualitativamente en relación a aspectos como la facilidad de consulta, la ayuda disponible, etc.

Palabras clave: Prensa. Ediciones electrónicas. Recuperación de información. Evaluación. Internet.

0.2 Abstract

Internet press editions are becoming more and more different from their traditional counterparts, due to the incorporation of new services that build on the possibilities of the new medium, mainly immediacy, world-wide dissemination, hypertext, multimedia, database integration, formatting flexibility, etc. Among such new services are the customisation of web access and e-mail news lists, and news retrieval engines. These last services are analysed quantitatively —using the information retrieval evaluation model — and qualitatively —examining aspects like friendliness.

Keywords: Press. Digital editions. Information retrieval. Evaluation. Internet.

1. Introducción

De entre los medios de comunicación de masas, fueron los impresos los primeros en ofrecer algún tipo de información en Internet. Al principio ésta no era más que un mero reflejo de la información aparecida en la edición impresa, imitando incluso el formato y aspecto de ésta. Poco a poco, fue creciendo la cantidad de información ofrecida, se fueron incorporando algunos servicios especializados y diversos elementos de valor añadido con los que trataban de aprovechar las posibilidades de Internet y algunas de sus principales características, tales como su inmediatez, posibilidad de difusión mundial, construcción de contenidos basados en hipertexto, multimedia, integración de bases de datos, flexibilidad de formatos, etc. De esta forma, fue haciéndose cada vez mayor el grado de independencia —de autonomía— de la edición electrónica respecto de la impresa. De esta forma, en la actualidad es posible observar que en determinados casos la edición impresa y la digital no solamente son ediciones distintas, sino que incluso forman parte de proyectos distintos perfectamente diferenciados. Por otro lado, también empiezan a proliferar medios que únicamente se pueden encontrar en la red y que hacen de ésta su única razón de ser.

Poco a poco, se van introduciendo en estos periódicos digitales nuevas herramientas y servicios, cuyo objetivo último es permitir un acceso lo más directo posible a la información que el lector precisa. En algunos casos, por ejemplo, se trata de sistemas de personalización y/o envío de noticias por correo electrónico, configurables por parte del usuario, o bien se aporta información multimedia (servicio de radio o videos de noticias) y, en la mayoría de los casos, el servicio que prácticamente todos los medios incorporan es un buscador de noticias, que permite y recuperar noticias disponibles en línea en ese medio, habitualmente los ejemplares de los anteriores 7-8 días, aunque algunos permiten buscar por meses completos —como *El Mundo*— o cubren un periodo de hasta 6 meses, como ocurre con *La Vanguardia*. Estos buscadores o sistemas de recuperación de información ofrecen distintas posibilidades de recuperación para cada medio, siendo posible, por ejemplo, en algunos casos, utilizar uno o más operadores booleanos para combinar términos, delimitar fechas u obtener resultados ponderados por relevancia.

2. Objetivo

En este texto se pretende hacer un pequeño estudio preliminar de los sistemas de recuperación de información de los principales medios de comunicación escrita en Internet. Más que una mera descripción de las características de cada uno de ellos se busca realizar una primera aproximación a la evaluación de sus sistemas de recuperación de información. Si bien es evidente la necesidad del estudio de datos derivados de la observación directa, con los que se pueda definir diversas características, parámetros e indicadores cualitativos, con los que eva-

luar aspectos tales como la facilidad de consulta, ayuda disponible, etc. No es esa, sin embargo, nuestra intención en este texto, pudiéndose encontrar suficientes textos en los que todas esas características se tratan con adecuada atención (Fuentes, 1998 ; Jiménez, González y Fuentes, 1999).

Se trata, por tanto, de hacer una primera aproximación a las posibilidades reales de recuperación de información que ofrece cada uno de dichos sistemas, así como tratar de valorar de la forma más objetiva posible su calidad basándose en la pertinencia de los resultados obtenidos tras la búsqueda, a partir de lo que se obtienen una serie de datos cuantitativos sobre la precisión y exhaustividad.

3. Metodología

Se ha partido de la adaptación y simplificación de algunas metodologías existentes para la evaluación de la recuperación de información —fundamentalmente en lo que a términos de exhaustividad y precisión se refiere—, concretamente respecto a los planteamientos metodológicos recogidos en los experimentos de Cranfield que propusiera Cleverdon (1972), posteriormente modificados por Salton y McGill (1983) (Hipola, Vargas-Quesada y Montes, 1999).

Elegimos cuatro diarios digitales españoles de carácter nacional (El País, ABC, El Mundo y La Vanguardia) sobre cuyo buscador de noticias lanzamos un total de diez consultas sobre diez temas de actualidad informativa general.

En un primer momento tratamos de emplear consultas lo más cercanas posible a la realidad, por lo que pedimos a usuarios no especializados que realizaran consultas en el buscador tal como las podrían hacer normalmente sin nuestra intervención, es decir, sin previamente haberles explicado detalladamente el funcionamiento de cada buscador. Sin embargo, dada la diferencia de funcionamiento de uno a otro buscador —de forma que unos usan por defecto el operador AND (*El Mundo*), otros el OR (*El País*) y otros la expresión literal (*ABC* y *La Vanguardia*)—, y dado que los usuarios no especializados tienden a no usar operador alguno, se llegó a una situación en la que una vez realizada una primera búsqueda, resultaron ser muy abundantes las consultas en las que no se recuperaba ni un solo documento, consecuencia probable de no haber utilizado adecuadamente el lenguaje de búsqueda del sistema. Este resultado nos hace poner en tela de juicio la utilidad real de estos buscadores para el usuario no especializado, para el que podría darse en un momento dado la situación de estar lanzando en un sistema búsquedas que, si bien serían perfectamente válidas en otro sistema que usara otros operadores de búsqueda, en este no son en absoluto efectivas.

Así, optamos finalmente por intervenir en el proceso de búsqueda, diseñando nosotros mismos las búsquedas, utilizando operadores cuando se hizo necesario.

Los temas de las consultas fueron diez: “Terrorismo”, “Mujeres Maltratadas”, “Ana Rosa Quintana”, “Bajada del IRPF”, “Plan Hidrológico Nacional”, “Conflicto Israel-Palestina”, “Milosevic”, “Elecciones USA”, “Subida de la gasolina” y “Adopciones de niños”. Para la elección de los temas de búsqueda fue preciso tener en cuenta que de los periódicos electrónicos estudiados no se disponía de un fondo documental excesivamente amplio, pues éstos conservan en la base de datos pública sólo los últimos artículos publicados, cubriendo un periodo que oscila entre una semana y seis meses. Por lo tanto, si queríamos obtener algún éxito en la recuperación de información resultaba necesario realizar consultas sobre temas de actualidad cubiertos en ese periodo de tiempo por los medios de comunicación estudiados. Así, finalmente, nos centramos en algunos de los temas de actualidad más tratados por la prensa nacional en el último tramo del mes de octubre de 2000. El hecho de contar sólo con un limitado número de noticias en el sistema nos hizo plantearnos hasta qué punto se cumpliría en este caso el problema de escala que plantearon tanto Resnikoff (1978) como Swanson (1977), quienes venían a decir que, si la muestra de la base de datos era demasiado pequeña no obtendríamos información fiable del rendimiento del sistema o de la adecuación de las estrategias utilizadas. Ésta habría de ser una de las hipótesis a confirmar posteriormente a partir del análisis de los resultados.

Una vez realizada la búsqueda se comprobaron todos y cada uno de los enlaces y se observó el contenido de los documentos para determinar la pertinencia o no del resultado. Si bien la pertinencia es un parámetro con un alto componente de juicio de valor y susceptible de ser medida con distintas gradaciones, optamos por determinar la pertinencia en términos absolutos, binarios, es decir, con dos únicos valores: “pertinente” o “no pertinente”.

Nº Doc.	Pertinente	Precisión	Exhaustividad
1	X	1/1=1	1/6=0,16
2	X	2/2=1	2/6=0,33
3	X	3/3=1	3/6=0,5
4		3/4=0,75	3/6=0,5
5	X	4/5=0,8	4/6=0,66
6	X	5/6=0,83	5/6=0,83
7		5/7=0,714	5/6=0,83
8		5/8=0,625	5/6=0,83
9	X	6/9=0,66	6/6=1
10		6/10=0,6	6/6=1

Tabla I. Ejemplo de tabla de datos para una consulta y diez documentos recuperados.

Se construyó una tabla (tabla 1), donde se incorporó este dato junto con el cálculo de la precisión y la exhaustividad. La exhaustividad para cada documento de la lista de documentos recuperados viene dada por el resultado de la división del número de documentos relevantes recuperados hasta ese momento entre el total de documentos relevantes recuperados para esa consulta ($E = drr / Tdrr$). Por su parte, la precisión para cada documento de la lista es el resultado de dividir el número de documentos relevantes recuperados hasta esa posición de la lista de recuperados entre la suma de los documentos relevantes recuperados y los documentos no relevantes recuperados hasta esa posición ($P = drr / (drr + dnrr)$).

Con estos datos se fue construyendo una matriz de resultados y se calculó la media de la matriz de resultados de exhaustividad y la correspondiente a la de precisión con el objeto de obtener una columna de datos que representara la exhaustividad y otra que representara la precisión, y así poder plasmar el resultado en una representación gráfica, a la que a continuación se le calculó la correspondiente curva de regresión logarítmica, con la que se pudo representar el funcionamiento general del ingenio de búsqueda (1) en cuestión.

El ingenio de búsqueda ideal sería aquel que recuperara únicamente documentos relevantes. Tal y como hemos calculado los valores de exhaustividad y precisión, ese sistema tendría asociada una representación gráfica consistente en una línea totalmente horizontal a la altura del máximo valor de precisión, es decir 1.

Sin embargo, los sistemas reales devuelven tanto documentos relevantes como no relevantes. Teniendo en cuenta este dato se considera que serán mejores sistemas aquellos que, pese a recuperar documentos no relevantes intercalados entre los relevantes, sean capaces, sin embargo, de devolvernos en los primeros lugares de la lista sobre todo documentos relevantes, y viceversa: que sitúen entre los últimos lugares de la lista sobre todo documentos no relevantes. Así, un sistema que ordene los resultados por orden de relevancia tenderá a ofrecer una curva de relación precisión/exhaustividad decreciente, y un sistema que no ordene por relevancia tenderá a ofrecer como resultado una línea plana por debajo del valor máximo de precisión.

Para comparar varios sistemas de recuperación de información a partir de su representación gráfica tendremos en cuenta ante todo el hecho de que para iguales valores de exhaustividad se obtengan mayores valores de precisión, así como el hecho de obtener una curva decreciente frente a una horizontal.

Dependiendo del periódico estudiado encontramos gran variabilidad en el número de resultados obtenidos para cada búsqueda. Por ello se consideró necesario tener en cuenta el número medio de documentos recuperados por consulta, y así hacer aparecer en la gráfica únicamente el número medio de resultados obtenido. Así se podían observar dos cualidades en la misma gráfica: por un lado la rela-

ción precisión/exhaustividad —que se habría de manifestar en la pendiente de la curva—, y, por otro, la profundidad del número de resultados que devuelve cada sistema de recuperación —lo que se observaría a partir de la longitud de la curva.

4. Resultados

El primer dato a considerar es la media de documentos recuperados por consulta en cada SRI, siendo el SRI de *La Vanguardia* el único que devolvió en todos los casos el máximo de respuestas tabuladas (20), seguido por el del *El Mundo* con 16, *El País* 15 y por último *ABC* con 8 documentos recuperados.

En cuanto a la relación entre exhaustividad y precisión obtuvimos como resultado cuatro matrices de datos, que representamos gráficamente mediante sus cuatro líneas de regresión correspondientes, con las que tratamos de mostrar visualmente el rendimiento de los ingenios de búsqueda.

Se observa fácilmente que el único sistema que muestra una relación claramente decreciente de precisión/exhaustividad es *El Mundo*, cuyo SRI ofrece una ordenación de resultados por grado de relevancia, representado éste por medio de porcentajes de adecuación a la query. *El País*, aunque también ofrece los resultados ordenados por relevancia (Jiménez, González y Fuentes, 1999) y usa en su sistema de recuperación el ingenio de búsqueda de Altavista Magallanes, a partir de los datos obtenidos experimentalmente no se muestra la relación de proporcionalidad inversa de precisión/exhaustividad de forma tan clara como lo hace el sistema de *El Mundo*, mostrando una línea de regresión con una inclinación bastante pequeña. Por su parte los SRI de *ABC* y *La Vanguardia* generan una línea prácticamente plana, lo que confirma el hecho de que estos sistemas no ofrecen ponderación en la lista de resultados.

También se muestra claramente que la relación precisión/exhaustividad del SRI de *El Mundo* es la mayor en prácticamente todos los tramos de la línea, lo que quiere decir que para cualquier valor de exhaustividad de este SRI siempre tendremos una mayor precisión que para el resto. Solamente al final de la línea de regresión de este sistema hay otro que se sitúa ligeramente por encima en cuanto a precisión (*La Vanguardia*), que muestra aproximadamente los mismos valores

Periódico	Nº medio de doc. Recup.
La Vanguardia	20
El Mundo	16
El País	15
ABC	8

Tabla II. Media de documentos recuperados

en todos los tramos de la línea, consecuencia, como dijimos anteriormente, de no ordenar sus resultados por algún tipo de orden de relevancia.

Si hiciéramos una valoración global podríamos decir con cierto margen de confianza que de los estudiados el mejor SRI, en cuanto a los indicadores evaluados (exhaustividad y precisión), es el de *El Mundo*. Esto será así muy probablemente por el hecho de que este sistema ordena los resultados por orden de relevancia, dejando siempre dentro de los límites estudiados (los veinte primeros resultados) los documentos más relevantes.

Exhaustividad/Precisión La Vanguardia

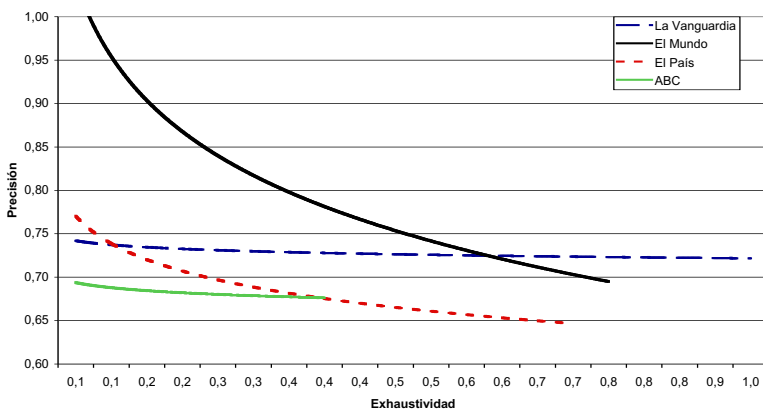
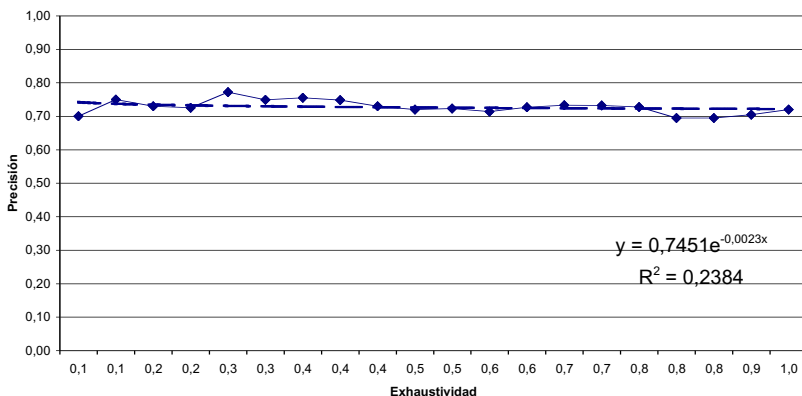


Figura 1. Ejemplo de dos gráficas de exhaustividad y precisión con su curva de regresión

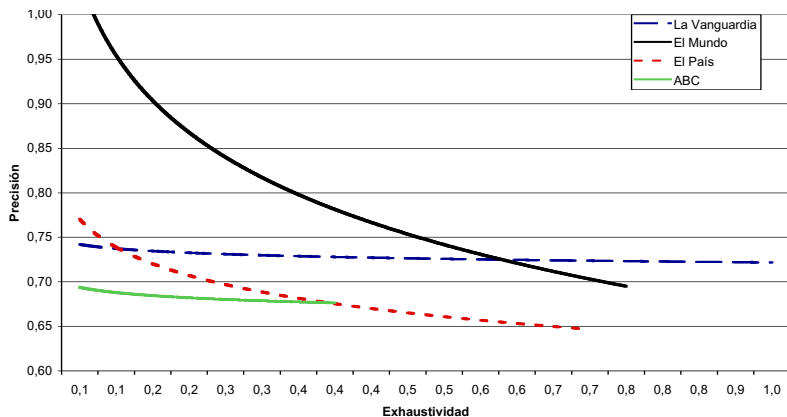


Figura 2. Relación exhaustividad/precisión en los SRI estudiados

Respecto a los otros tres sistemas, *La Vanguardia* obtiene mejores resultados que los otros dos, posiblemente por el hecho de que es el que mayor base documental contiene de todos los SRI estudiados (últimos seis meses).

5. Conclusiones

- El hecho de que el SRI sea capaz de ordenar los resultados por orden de relevancia será un factor clave para conseguir obtener mejor relación precisión/exhaustividad que aquellos que no lo hacen, aún cuando los que ordenan los resultados cuenten en su base documental con un número sustancialmente menor de documentos.
- El fracaso en el intento de búsqueda por parte de usuarios reales con el que nos tuvimos que enfrentar nos dice poco de las interfaces de búsqueda, pues en muchos casos una búsqueda sin operadores se traducían directamente en una búsqueda totalmente incorrecta para el sistema y devolvía 0 resultados. Si no se mejora la interfaz y se incorpora una pequeña ayuda directamente junto a la ventana de búsqueda, o se impide que el operador por defecto sea la cadena literal (por ejemplo, usando subida gasolina nos devolvería probablemente 0 documentos al considerarla como expresión literal) lo que tiende a frustrar a usuario sin que llegue a darse cuenta de cual es el verdadero motivo por el que no obtiene resultado alguno.
- Tal como indicaban Resnikoff (1978) y Swanson (1977) el tamaño de la base de datos es demasiado pequeño como para tener un fondo suficiente-

mente abundante sobre el que experimentar, en el que se pudiera lanzar búsquedas en las que el mayor o menor éxito de la recuperación dependiera únicamente de la consulta y del ingenio de búsqueda y no de la adecuación y cobertura del propio fondo documental. Esto nos hace plantearnos que el experimento tal como lo hemos planteado no parece ofrecer toda la luz que debiera acerca de la calidad de los ingenios de búsqueda de estos diarios.

- Por ello, teniendo en cuenta el limitado tamaño de la base de datos, sería interesante replantear el experimento teniendo en cuenta la totalidad de los documentos de la base de datos, es decir, contabilizar todos los relevantes respecto del total de base de datos y no del total de recuperados, así como no limitar el estudio a los *x* primeros documentos recuperados.
- Otro aspecto cuestionable es la *formulación de las consultas*. Tras intentar dejar la realización de las preguntas en manos de los usuarios, se hizo necesario replantear las consultas, lo que quiere decir que si no se efectúa una consulta con cierto grado de perfección no se obtendrán apenas resultados. Es decir, la propia formulación de las consultas es un elemento clave y tremendamente susceptible de interpretación y, por lo tanto, difícilmente objetivable. Se podría recurrir a usar un grupo de expertos para definir las consultas más adecuadas, pero ello significaría que estaríamos creando unas condiciones irrealistas. Nos interesa ser capaces de evaluar el sistema de cara a los usuarios, no de cara a los expertos. De poco sirve que un equipo de expertos sea capaz de formular una pregunta adecuada y evaluar la pertinencia de los resultados si un usuario medio no va a poder llegar a conclusiones similares y por tanto, obtener resultados similares.

6. Notas

- (1) Si bien en la bibliografía existente se tiende a traducir *search engine* por motor de búsqueda, nosotros nos inclinaremos por ingenio de búsqueda por parecernos una traducción más adecuada o bien utilizaremos simplemente SRI (Sistema de Recuperación de Información).

7. Bibliografía

- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. // *Journal of Documentation*. 56 : 1 (Jan. 2000) 71-90.
- Cleverdon, Cyril W. (1972). On the inverse relationship of recall and precision. // *Journal of documentation*. 28 (1972) 195-201.
- Fuentes i Pujol, M. E. ; González Quesada, A. (1998). La prensa española en Internet: análisis de los servicios de valor añadido. // VI Jornadas Españolas de Documentación. Valencia : FESABID, 1998. 281-292.

- Harter, Stephen P. ; Hert, Caroll A. (1997). Evaluation of information retrieval systems: approaches, issues and methods. // *Annual review of information science and technology*. 32 (1997) 3-94.
- Hipola, P. ; Vargas-Quesada, B. ; Montes, A. (1999). Descripción y evaluación de agentes multibuscadores. // *El profesional de la información*. 8 : 11 (noviembre, 1999)
- Jiménez López, A. ; González Quesada, A.; Fuentes i Pujol, M .E. (1999). Gestió documental de la informació en els serveis de valor afegit de la premsa espanyola a Internet [en línea]. // VII Jornades Catalanes de Documentació. Barcelona, 1999. URL: <<http://www.cobdc.org/7es/18.pdf>>. Consultado: 6/2/2000.
- Oppenheim, C. ; Morris, C. ; McKnight (2000). The Evaluation of WWW search engines. // *Journal of Documentation*. 56 : 2 (March, 2000) 190-211.
- Resnikoff, H. L. (1978). The National Need for Reseach in Information Science. // STI Issues an Options Workshop, House subcommittee on science, research an technology. Washington, D. C. Noviembre, 3, 1978.
- Salton, G. ; McGill, J. (1983). Introduction to modern information retrieval. New York : McGraw-Hill, 1983.
- Swanson, D. R.(1977). Information Retrieval as a Trial an Error Process. // *Librari Quaterli*. 47 : 2 (1977) 128-148.
- Wang, H. ; Xie, M.; Goh, T. N. (1999). Service quality of Internet search engines. // *Journal of Information Science*. 25 : 6 (1999) 499-507.