

# Artículo Interacción entre medidas de popularidad en el posicionamiento web

Por Valentín Moreno Pelayo

**Resumen:** El trabajo se centra en la relación entre algunas medidas de popularidad y posicionamiento de las páginas recuperadas en los buscadores Google y MSN Search. Se ha analizado la interacción entre el número de enlaces a una página y su tráfico. Se ha obtenido como conclusión que en Google, de forma más decisiva que en MSN Search, el posicionamiento de un resultado está determinado por el número de enlaces en combinación con el tráfico. Google posiciona mejor las páginas con un elevado número de enlaces y que, al mismo tiempo, tienen un tráfico acorde a los enlaces que recibe. Esta estrategia permite luchar a Google contra la contaminación, es decir páginas que reciben muchos enlaces pero pocas visitas.

**Palabras clave:** PageRank, Visibilidad, Algoritmo de posicionamiento, Alcance, Tráfico.



Valentín Moreno Pelayo es matemático (especialidad computación), profesor del Departamento de Informática de la Universidad Carlos III de Madrid, investigador en el proyecto «Sistema avanzado de asistencia a la conducción para entornos urbanos: inteligencia artificial» financiado por Cicyt (2004-07), TRA2004-07441-C03-02/AUT, colaborador en investigación sobre redes temáticas.

**Title:** Interaction of popularity measurements in web positioning

**Abstract:** This work focuses on the relationship between certain popularity measurements and the positioning algorithm of pages retrieved by Google and MSN Search browsers. Measurements studied were: number of page links, number of visitors («reach»), and number of URL requests on a site («pageviews»). The combination of the last two parameters determines page traffic. The following conclusions were drawn from this research: in Google, location of a result, as determined by the number of links, is supported by the page traffic; thus Google performs better in locating those pages in which a high number of links correlates with a high page traffic. This enables Google to avoid contamination: that is, pages with a high number of links and a low number of visits.

**Keywords:** PageRank, Visibility, Positioning algorithm, Reach, Traffic rank.

**Moreno Pelayo, Valentín.** «Interacción entre medidas de popularidad en el posicionamiento web». En: *El profesional de la información*, 2005, marzo-abril, v. 14, n. 2, pp. 100-107.

## Introducción

La optimización de páginas web para aumentar su visibilidad es un tema clave para los creadores de páginas web<sup>1</sup>. La técnica consiste básicamente en conocer los criterios que utiliza el algoritmo de posicionamiento de cada buscador para mejorar la posición de la página. Uno de los criterios que más se viene utilizando es la medida de la popularidad de determinada página y Google<sup>2</sup>, con el PageRank<sup>3</sup>, diseñó un método para medirla basado en el número de enlaces entrantes a la página. Lamentablemente, los webmasters han restado eficacia al aumentar la contaminación mediante la utilización de políticas de intercambios y “granjas de enlaces”. Otra forma de medir este dato es calcular el tráfico de visitas que tiene una página (p. e. Alexa). En este artículo se estudian posibles relaciones entre:

—Enlaces a la página.

—Cantidad de visitantes.

—Número de peticiones de urls en determinado sitio.

Por lo tanto, se tratará de analizar la relación entre el número de enlaces y el tráfico de determinada página, indicado por los dos últimos factores. Es lógico pensar que a mayor número de enlaces apuntando habrá un mayor número de visitas, por lo que tráfico y enlaces no son variables independientes. En la práctica, las estrategias de optimización hacen que se falsee el número de links, pero la cantidad de visitas es más complicada de manipular. De hecho, un análisis de las páginas indica que la proporción en que se presentan en cada una es en general diferente (a casos con el mismo número de enlaces no les suele corresponder el mismo tráfico y viceversa). Se estudiará si esta rela-

Artículo recibido el 25-09-04  
Aceptación definitiva: 11-02-05

ción depende del buscador y cómo afecta al posicionamiento.

En las siguientes secciones se definen, en primer lugar, factores relevantes para el estudio como son el “alcance” y las “páginas visitadas”. Posteriormente se realizará un análisis de estos factores basado en los coeficientes de correlación, y por último ofreceremos una posible explicación de los datos observados.

## Método

En el presente apartado se describen las siguientes fases:

—Obtención de datos.

—Cálculo de diferentes coeficientes de correlación.

—Comprensión del algoritmo de posicionamiento.

### 1. Obtención de datos

El análisis se centra en la popularidad. Al no disponer directamente de la información que tienen los buscadores sobre el tráfico, es necesaria una herramienta que aporte datos estadísticos fiables para su estudio. En este caso la aplicación para la obtención de esa información ha sido *Alexa*<sup>4</sup>, selección motivada por su fiabilidad y precisión, aunque posee ciertas limitaciones que se comentan a continuación.

*Alexa* es un sistema de evaluación utilizado y aceptado como parámetro de referencia en el ranking

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	21.905	6	3.384	2
2	22.299	6	767	3
3	27.903	6	214	4
4	308	3	174	4
5	30.334	6	208	4
6	33.669	6	46	5
7	6.060	5	1.066	3
8	33	2	11.706	2
9	12.201	5	39	5
10	99.076	7	1.387	3
11	291	3	1.622	3
12	70.519	6	3	7
13	5.883	5	20	6
14	28.291	6	3	7
15	2.397	4	367	4

Tabla 1. Consulta en Google “adn”

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	6.060	5	1.066	3
2	38.288	6	5	6
3	26.466	6	14	6
4	48.668	6	11	6
5	308	3	174	4
6	291	3	36	5
7	326	3	1	7
8	1.093	4	6.396	2
9	20.164	6	187	4
10	30.334	6	208	4
11	70.519	6	3	7
12	41.689	6	44	5
13	2.251	4	1	7
14	77.399	6	100	5
15	1.311	4	1.831	3
16	5.539	5	294	4
17	2	0	28	5

Tabla 2. Consulta en MSN Search “adn”

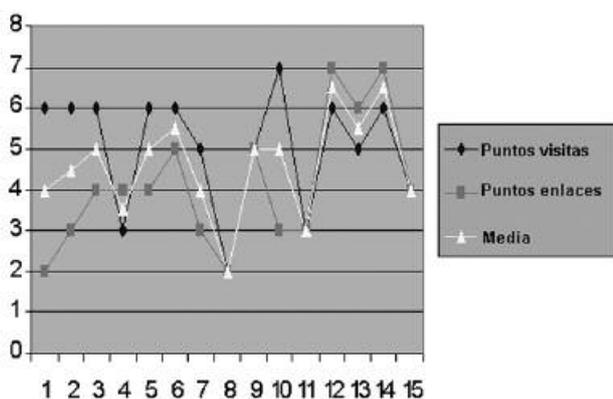


Imagen 1. Gráfica comparativa tráfico/enlaces Google (adn)

de “popularidad” por los grandes sitios y las grandes empresas en internet. Maneja información sobre sitios web relacionados, estadísticas de visitas, valoración de los usuarios, propietarios, fecha de creación y además realiza comparativas de tráfico con otros sitios analizando semanalmente las tendencias en visitas y páginas vistas.

#### a. Ranking de *Alexa*.

a.1. Basado en las visitas de los internautas que tienen instalada su barra (más de 10 millones en todo el mundo) en períodos de tres meses.

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	12.071	5	328	4
2	58.031	6	558	4
3	5.530	5	173	4
4	11.653	5	18	6
5	11.653	5	21	6
6	10.206	5	2.538	3
7	56.663	6	59	5
8	51.110	6	123	5
9	1.696	4	8.938	2
10	38.405	6	61	5
11	2.082	4	428	4
12	11.653	5	24	6
13	3.732	5	689	3
14	16.706	6	124	5
15	1.579	4	10	6
16	52.865	6	320	4
17	52.522	6	614	4
18	3	0	4.423	2
19	5.530	5	173	4
20	48.340	6	7	6
21	23	1	48.562	1
22	1.737	4	2.786	3
23	3.956	5	6.690	2
24	2.957	4	127	4

Tabla 3. Consulta en Google "information retrieval"

a.2. La posición que ocupa un sitio en el ranking mundial es una combinación del alcance y páginas vistas obtenidas, definiéndose estos parámetros como:

—Alcance ("reach"): número de usuarios (direcciones IP) que visitan un sitio en un día dado.

—Páginas visitadas ("page views"): cantidad de páginas visitadas por las urls diferentes que visitan un

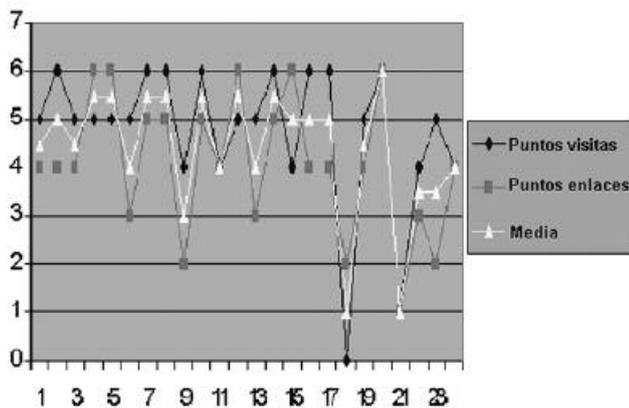


Imagen 3. Gráfica comparativa tráfico-enlaces Google (information retrieval)

sitio. En distintos días la misma url se cuenta como diferente.

b. Sesgos Alexa.

b.1. Funcionamiento limitado a los webs de nivel superior (del tipo "www.dominio.com").

b.2. Sólo funciona con el navegador Internet Explorer y el sistema operativo Windows.

b.3. Los factores culturales y la lengua (la información está en inglés) influyen en la adopción de su software.

b.4. Se desactiva en las páginas seguras (https:) de los sitios.

b.5. Los sitios con una posición por encima del puesto 100.000 no son fiables (con menos de 1.000 visitantes mensuales), ya que la cantidad de datos obtenida no es estadísticamente significativa.

## 2. Desarrollo y resultados

En estudios previos se observó la correlación entre el número de enlaces entrantes y las visitas en la posición de determinada página. Para comprobar este hecho se han realizado bús-

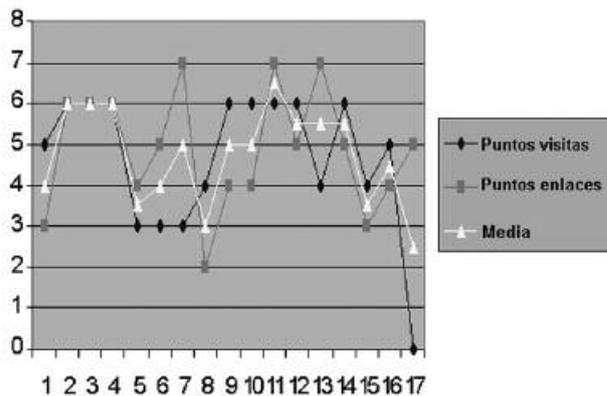


Imagen 2. Gráfica comparativa tráfico-enlaces MSN Search (adn)

quedas en Google y MSN Search. Con un enfoque análogo al propuesto por Chignell<sup>5</sup> (1999) se han estudiado los cuarenta y cinco primeros resultados recuperados, frente a los veinte propuestos por este autor, y se han analizado los factores relativos a enlaces entrantes y tráfico.

Para este artículo se han seleccionado cinco consultas de entre setenta y cinco realizadas con resultados similares: "adn", "information retrieval", "tutorial sql", "BOE vivienda", "tutorial xml". A partir de estas

búsquedas se han analizado más de ochenta páginas en lo relativo a los factores reseñados anteriormente.

## 2.1. Fases del desarrollo

### 2.1.1. Obtención de datos

I. Número de resultados por búsqueda y buscador

a. Adn:

—*Google*: 2.600.000.

—*MSN Search*: 1.392.424.

b. Information retrieval:

—*Google*: 6.820.000.

—*MSN Search*: 6.584.613.

c. Tutorial sql:

—*Google*: 2.460.000.

—*MSN Search*: 2.336.830.

d. BOE vivienda:

—*Google*: 125.000.

—*MSN Search*: 25.580.

e. Tutorial xml:

—*Google*: 5.080.000.

—*MSN Search*: 2.755.885.

II. Para cada resultado devuelto en cada búsqueda obtenemos los siguientes datos:

a. Posición entre los resultados de la búsqueda.

b. Número de enlaces a la página.

c. Ranking *Alexa* (posición por tráfico obtenida combinando los criterios siguientes):

—Número de visitantes (“alcance”) en tres meses (medidos en millones).

—Cantidad de páginas visitadas (desde ese resultado).

### 2.1.2. Comprensión del algoritmo de posicionamiento

Para entender cómo afecta la relación entre el tráfico y el número de enlaces de una página en el posicionamiento es conveniente recordar que se puede hacer una aproximación a *Google* para puntuar las páginas en fun-

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	12.071	5	328	4
2	58.031	6	558	4
3	5.530	5	173	4
4	11.653	5	18	6
5	377	5	10.682	2
6	11.653	5	18	6
7	10.206	5	2.538	3
8	2.902	4	53	5
9	23	1	48.562	1
10	5.530	5	173	4
11	12.071	5	50	5
12	16.706	6	47	5
13	38.405	6	61	5
14	1.864	4	105	5
15	1.239	4	871	3
16	1.696	4	8.938	2
17	1.356	4	144	4
18	8.161	5	22	6
19	4.189	5	3.581	2
20	58.031	6	558	4
21	2.957	4	668	3
22	377	3	10.682	2
23	51.110	6	123	5
24	1.297	4	25	5
25	2.082	4	428	4
26	2.957	4	396	4

Tabla 4. Consulta en *MSN Search* “information retrieval”

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	1.930	4	1.447	3
2	77.087	6	39	5
3	4.401	5	23	6
4	11.884	5	4.858	2
5	3.386	5	20	6
6	65.854	6	25	5
7	629	4	6	6
8	172	3	16.142	1
9	2.175	4	283	4
10	15.652	6	380	4
11	133	3	318	4
12	66.202	6	21	6
13	52.933	6	88	5
14	25.704	6	3	7

Tabla 5. Consulta en *Google* “tutorial sql”

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	1.930	4	1.447	3
2	77.087	6	39	5
3	4.401	5	23	6
4	11.884	5	4.858	2
5	94.785	7	21	6
6	65.854	6	25	5
7	133	3	318	4
8	1.489	4	6	6
9	1.930	4	1.447	3
10	4.401	5	23	6
11	77.087	6	39	5
12	3.386	5	20	6
13	2.175	4	283	4
14	126	3	11	6
15	2.306	4	8	6
16	9.086	5	1	7
17	7.849	5	168	4
18	15.036	5	386	4
19	59.036	6	244	4
20	10.757	5	41	5
21	22.506	6	40	5
22	14.195	5	335	4

Tabla 6. Consulta en MSN Search "tuturial sql"

Puntos	N. de enlaces
1	5
2	25
3	125
4	625
5	3125
6	15625
7	más de 15625

Tabla para la comprensión del algoritmo de posicionamiento

ción del número de enlaces que recibe mediante una escala logarítmica (puede verse un ejemplo en la tabla adjunta dedicada a este punto). Actualmente son al menos diez categorías, y posiblemente se aproxime con un logaritmo distinto del  $\log_5$ , aunque cualitativamente funciona de modo similar.

**«Uno de los criterios que más se viene utilizando es la medida de la popularidad de determinada página y Google, con el PageRank, diseñó un método para medirla basado en el número de enlaces entrantes a la página»**

Basándose en esta idea y con el fin de comparar gráficamente los datos de enlaces con los del tráfico, se aplicó el  $\log_5$  a ambos unificando así la escala en que van a ser representados. Además, sendas gráficas deben de tener el mismo criterio de ordenación respecto al posicionamiento.

Las etapas que se han seguido son:

—Seleccionar sólo las páginas con ranking *Alexa* menor que 100.000 (ver los sesgos de *Alexa*). Cabe destacar que en la mayoría de las búsquedas realizadas las pági-

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	47	2	163	4
2	97.488	7	46	5
3	5.087	5	465	3
4	5.381	5	195	4
5	9.718	6	250	4
6	6.783	5	334	3
7	4.828	5	233	4
8	39.086	7	67	4
9	19.699	6	35	5
10	23.070	6	67	4
11	43.864	7	257	4
12	15.989	6	251	4
13	16.372	6	149	4
14	3.339	5	58	4
15	27.932	6	39	5
16	11.179	6	291	3
17	7.856	6	349	3
18	38.851	7	152	4
19	6.021	5	25	5
20	16.372	6	149	4
21	13.624	6	98	4
22	86.375	7	28	5

Tabla 7. Consulta en Google "BOE vivienda"

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	97.488	7	46	5
2	5.087	5	465	3
3	6.783	5	334	3
4	639	4	8	6
5	61.857	7	136	4
6	2.153	5	1.169	3
7	19.516	6	137	4
8	6.783	5	18	5
9	63.206	7	34	5
10	5.087	5	3	6
11	1.929	5	324	3
12	29.035	6	198	4
13	9.718	6	250	4
14	7.856	6	114	4
15	39.086	7	67	4

Tabla 8. Consulta en MSN Search "BOE vivienda"

nas que cumplen este requisito son casi la mitad, una muestra suficientemente significativa ya que partíamos de cuarenta y cinco resultados.

—Obtener la información sobre el número de enlaces y el ranking *Alexa* (tráfico) y aplicarles  $\log_5$ . Los nuevos datos se denominarán puntos enlaces y puntos tráfico respectivamente.

—Cuanto menor sea el valor de puntos tráfico mejor posición, siendo al contrario para puntos enlaces. Reemplazando puntos enlaces por siete menos puntos enlaces invertimos el criterio de ordenación facilitando así el estudio de la relación entre ambos criterios.

—Construir gráficas comparativas para puntos enlaces y puntos tráfico por búsqueda y buscador.

### Conclusiones

Aunque es lógico pensar que a mayor número de enlaces más visitas, el estudio muestra evidencia de que la correlación entre uno y otro no es perfecta. También se han encontrado diferencias entre *Google* y *MSN Search*. En el primero el posicionamiento de un resultado determinado por el número de enlaces es avalado por el tráfico en el mismo. El coeficiente de corre-

lación está próximo a 0,5 de media, suficientemente significativo si consideramos la gran cantidad de factores que influyen en el posicionamiento.

Se observa que las gráficas realizadas con los datos de los puntos por enlace y los puntos por tráfico son cualitativamente similares, es decir, *Google* posiciona mejor la página cuyo número de enlaces a la misma es elevado y además acorde al tráfico que recibe. Esto le permite luchar contra la contaminación (las páginas que reciben muchos enlaces pero pocas visitas no se van a posicionar bien). Es sencillo incluir esta idea en una fórmula: es suficiente con obtener la posición de una página considerando el criterio del número de enlaces, conseguir su posición atendiendo ahora al criterio de tráfico recibido y por último calcular la distancia entre ambas. Si esta diferencia es pequeña su posicionamiento se verá favorecido.

*MSN Search* también utiliza estos dos criterios para posicionar pero sin exigir concordancia entre ellos

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	1.943	5	1.447	2
2	721	4	5.946	2
3	78.874	7	760	3
4	21.084	6	1.333	3
5	56.116	7	440	3
6	10	1	4.868	2
7	721	4	5.946	2
8	9.596	6	1.326	3
9	39.494	7	946	3
10	50.719	7	132	4
11	39.494	7	946	3
12	1.473	5	687	3
13	10.770	6	468	3
14	333	4	7.822	1
15	64.074	7	498	3
16	333	4	7.822	1
17	64.074	7	498	3
18	89.993	7	386	3
19	2.356	5	1.186	3
20	1.479	5	10.686	1
21	8.886	6	2.360	2
22	8.638	6	584	3
23	22.349	6	1.507	2

Tabla 9. Consulta en Google "tutorial xml"

Posición	Ranking Alexa	Puntos por visitas	Enlaces a la página	Puntos enlaces
1	1.943	5	1.447	2
2	721	4	5.946	2
3	10	1	4.868	2
4	56.116	7	440	3
5	21.084	6	1.333	3
6	1.479	5	10.639	1
7	21.680	6	921	3
8	89.993	7	386	3
9	8.638	6	584	3
10	1.463	5	6	6
11	65.611	7	965	3
12	9.887	6	1.109	3
13	18.454	6	953	3
14	25.926	6	98	4
15	63.004	7	104	4
16	1.070	4	432	3
17	89.993	7	386	3

Tabla 10. Consulta en MSN Search "tutorial xml"

(en varias búsquedas el coeficiente de correlación está muy próximo a cero). En las gráficas se observa que posiciona bien las páginas con muchos enlaces y en proporción menos tráfico, y viceversa. Si una página destaca en uno de los criterios puede compensar un resultado modesto en el otro.

**«En la práctica, las estrategias de optimización hacen que se falsee el número de links, pero la cantidad de visitas es más complicada de manipular»**

Así, se puede concluir que aunque Google da más importancia a los enlaces entrantes, algo corroborable por la

	Adn	Information retrieval	Tutorial sql	BOE vivienda	Tutorial xml
Google	0,307	0,589	0,539	0,30	0,768
MSN Search	0,081	0,521	0,135	-0,013	0,345

Tabla 11. Coeficientes de correlación (tráfico/enlaces)

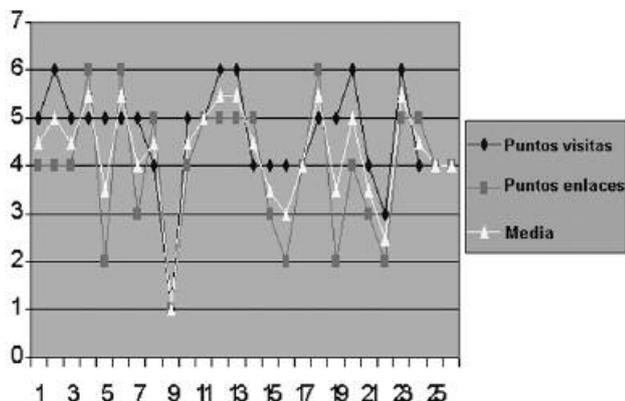


Imagen 4. Gráfica comparativa tráfico-enlaces MSN Search (information retrieval)

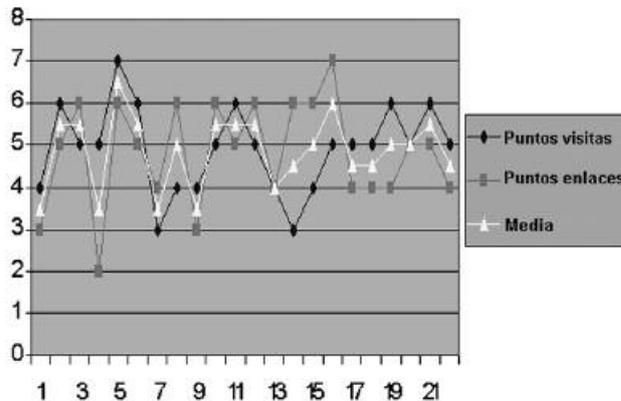


Imagen 6. Gráfica comparativa tráfico-enlaces MSN Search (tutorial sql)

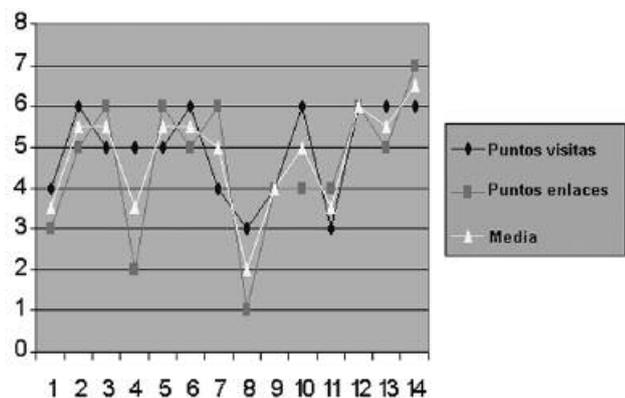


Imagen 5. Gráfica comparativa tráfico-enlaces Google (tutorial sql)

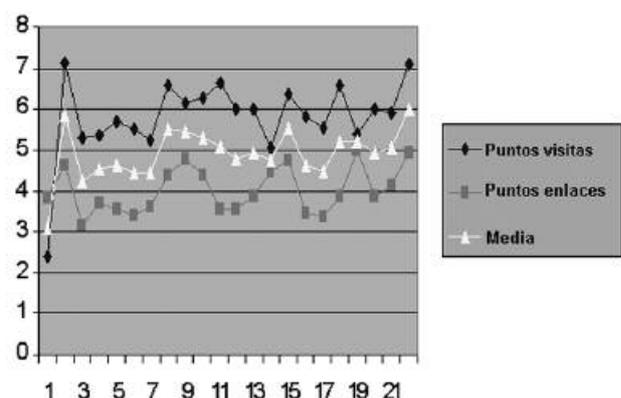


Imagen 7. Gráfica comparativa tráfico-enlaces Google (BOE vivienda)

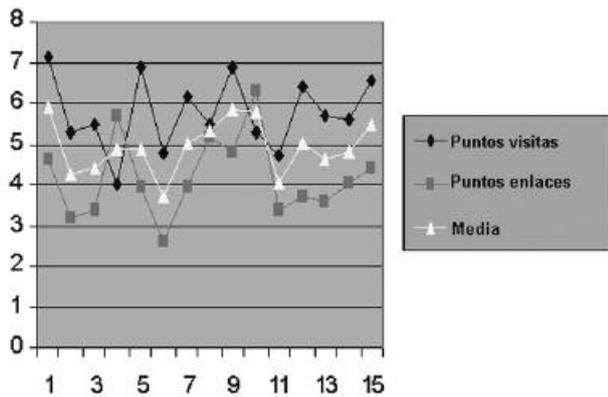


Imagen 8. Gráfica comparativa tráfico-enlaces MSN Search (BOE vivienda)

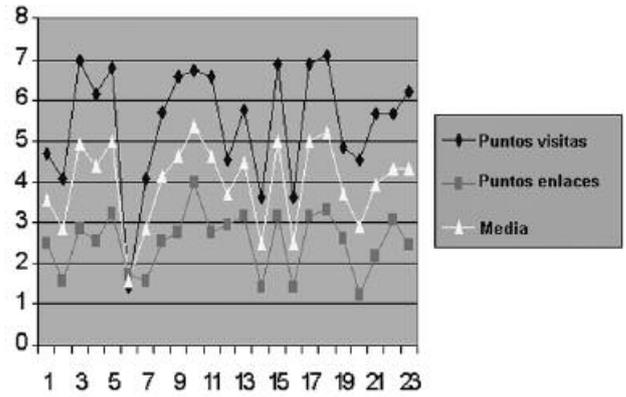


Imagen 9. Gráfica comparativa tráfico-enlaces Google (Tutorial xml)

estabilidad de sus resultados, debe estar teniendo en cuenta de alguna forma el número de visitas. Esta circunstancia no se da en *MSN Search*, lo que creemos que es algo consciente y no una consecuencia indirecta de la dependencia entre tráfico y enlaces ni debido a que contabiliza de igual forma los vínculos entrantes de páginas mediocres y los de elevada calidad.

Otras observaciones relevantes extraídas son:

—La relación entre el alcance y las páginas visitadas y tráfico es muy significativa: coeficientes de correlación próximos a 1, en torno a 0,998.

—*Google* recupera un número significativamente mayor de documentos (en algunas búsquedas casi el doble).

—En el cálculo del tráfico se da casi el doble de peso al alcance que a las páginas visitadas (según lo realiza *Alexa*).

—*MSN Search* y *Google* recuperan en las primeras posiciones un número similar de documentos por debajo de la posición 100.000 de la clasificación por tráfico que hace *Alexa*.

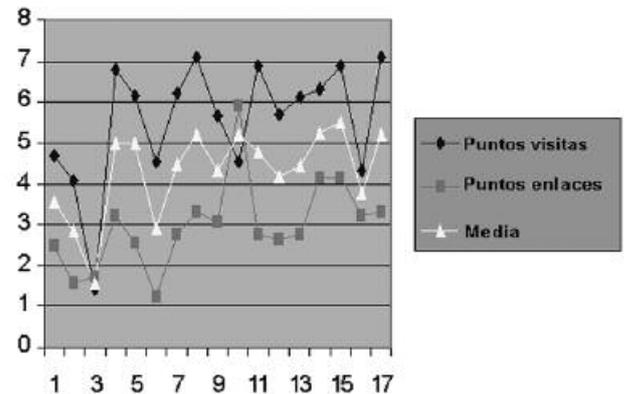


Imagen 10. Gráfica comparativa tráfico-enlaces MSN Search (tutorial xml)

transactions on knowledge and data engineering, 2004, January, v. 16, n. 1, pp. 4-16.

2. Google. *Google PageRank, GoogleMania, Todo sobre Google*. Consultado en: 01-10-04. <http://www.googlemania.com/pagerank.php>

3. 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.

4. Alexa Internet, Inc. *Alexa Web Search. About the Alexa traffic rankings*. Consultado en: 01-10-04. [http://pages.alexa.com/prod\\_serv/traffic\\_learn\\_more.html](http://pages.alexa.com/prod_serv/traffic_learn_more.html)

5. 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.

Valentín Moreno Pelayo, Departamento de Informática, Universidad Carlos III de Madrid. [vmpelayo@inf.uc3m.es](mailto:vmpelayo@inf.uc3m.es)

### Bibliografía

1. Diligenti, Michelangelo; Fellow, Marco Gori; Maggini, Marco. "A unified probabilistic framework for web page scoring systems". En: *IEEE*

### 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.elprofesionaldelainformacion.com/contenidos.html>