

# Análisis de los principales factores para el posicionamiento orgánico en Google.

Victor Estrugo Rottenstein, Facultad de Ingeniería,  
*Universidad de Palermo, Buenos Aires, Argentina*  
 <victor.a.rottenstein@gmail.com>

**Abstract—** Dado que las empresas buscan posicionarse entre los primeros resultados de búsqueda en Google, este trabajo pretende determinar si existe la posibilidad de definir factores para el posicionamiento de dichos resultados. Los resultados aquí presentados muestran que, si bien es posible establecer una incidencia, Existiría un condicionamiento previo (no relacionado con los estudiados aquí) que determinaría la inclusión en los tres primeros decálogos.

## I. KEYWORDS

Internet, Information retrieval, Optimization methods.

## II. INTRODUCCIÓN

Desde hace algún tiempo atrás los buscadores comenzaron a tomar un rol protagónico [14] en la relación entre los consumidores y las empresas a causa de que se transformaron en la interfaz mediante la cual un usuario accede al contenido distribuido por Internet, ya que es una práctica común el hecho de utilizar la red por medio de un buscador, aún cuando se recuerde el nombre del sitio al cual se desea acceder.

Se han realizado diversos estudios que determinan que el mayor porcentaje de personas que realizan búsquedas por Internet, solo acceden a los primeros 10 o 20 resultados de búsqueda [2]. Como consecuencia, y según la popularidad de cada búsqueda estos 20 primeros sitios generan tráfico calificado al sitio.

En este contexto las empresas comenzaron a analizar el valor que tiene para la compañía aparecer entre los primeros resultados de una búsqueda dada, ya que los beneficios tangibles comienzan con tráfico calificado pero podrían transformarse en oportunidades comerciales.

Con este escenario e impulsado por el Marketing nació el SEO (Search engine Optimization) creando una disciplina para la Web focalizada en optimizar los sitios para motores de búsqueda. Se trata de realizar distintas actividades, teniendo en cuenta varios factores para lograr posicionar un sitio entre los primeros resultados orgánicos para uno o varios términos de búsqueda.

Cuando se habla de SEO se estudian estos “factores” que son influyentes a la hora de posicionar un sitio, pero lo cierto es que aunque se contraste empíricamente cada uno de los mismos, solo serían adecuados para el determinado

espacio/tiempo en los que sean justificados, ya que aunque se puedan unir todos los factores existentes y llegar a conclusiones científicas, es posible que un instante después la realidad se vea afectada por una simple modificación en un algoritmo.

La presente investigación se desarrolla de la siguiente forma: se plantean los factores más importantes que influyen en el posicionamiento orgánico. [17]. Cada uno de estos factores se toman como una hipótesis distinta sobre el posicionamiento en buscadores, y se ponen a prueba mediante experimentación, el resultado luego de la contrastación de cada factor según el método inductivo, determinará para cada hipótesis, si es “corroborada” (serán ciertos hasta que alguien pruebe lo contrario), o si es refutada. (Con una sola hipótesis falsada bastaría para descartarlo definitivamente.).

Por otra parte, quizás no tiene sentido realizar una investigación sobre todos los factores necesarios para optimizar un sitio, ya que es probable que sea un trabajo infinito porque los algoritmos de ordenamiento son propietarios y por la problemática anteriormente planteada acerca de la temporalidad de las contrastaciones. Sin embargo, sería interesante centrarse en algunos factores considerando los papers utilizados como referencias en la presente investigación, por ejemplo, es sabido que para Google es realmente importante la estructura de los links para ordenar los resultados de búsqueda [1], entonces, los experimentos deberán focalizarse en los factores que se puedan contrastar empíricamente sin necesidad de establecer hipótesis ad-hoc.

Hasta ahora se han encontrado diversas fuentes sobre el SEO como herramienta de marketing [2], pero difícilmente se ha encontrado alguna investigación pública con rigor científico que determine los factores que influyen al posicionamiento en los principales buscadores. Entonces es dable de preguntarse lo siguiente:

¿Hasta que punto los factores existentes para el posicionamiento tienen rigor científico?, ¿Porque realmente nadie puede garantizar llegar 1ro en buscadores?, **¿Existe acaso una forma “científica” para posicionar orgánicamente un sitio?**, ¿Es posible hacer ingeniería inversa para comprender como funcionan los algoritmos de búsqueda y relevancia?

No se conoce cada una de las respuestas a estas preguntas, pero los resultados de este paper son un modesto aporte para descubrir algunas bases científicas sobre el posicionamiento en buscadores.

### III. ANÁLISIS DE LOS FACTORES

No es el objetivo de este trabajo corroborar o refutar todos los factores que influyen en el posicionamiento orgánico, sino, poder establecer empíricamente si los principales factores [17] tienen influencia sobre el posicionamiento, tomando a los mismos como hipótesis para el experimento.

Para realizar el experimento se toman algunos de los factores del posicionamiento orgánico en Google postulados en la encuesta realizada por SeoMoz [17] donde se exponen los que serían los 10 factores más relevantes para posicionar un sitio en buscadores en el año 2007: Uso de keyword en el Title, Popularidad (enlaces) global de una Web, Anchor Text de los backlinks, Estructura interna de los enlaces, Edad de una Web, Relación temática de las Web/páginas que enlazan respecto a una keyword, Popularidad (enlaces) de una página dentro de la comunidad temática, Uso de keywords en la página, Popularidad (enlaces) globales de página/Web que enlaza, Relación temática de las páginas que enlazan.

Hay factores que no se considerarán por los siguientes motivos:

- Falta de las herramientas de software adecuadas.
- No se pretende hacer ingeniería inversa de todos los algoritmos de los principales buscadores, sino una comprobación estadística del comportamiento en función de los factores.

Por otra parte, tampoco es objetivo de este trabajo comprender todos los factores según la encuesta de referencia, ya que sólo representa la opinión de expertos en el tema, y no de un análisis riguroso basado en experimentación.

A continuación se presenta cada uno de los factores que se analizaron en esta investigación, junto con una explicación de la forma y/o unidad mediante la cual se han utilizado.

1. **Uso de keyword en el Title:** Colocar el término de búsqueda o la frase en el tag llamado título del encabezado de la página HTML. (valor binario, si/no)
2. **Anchor Text de los backlinks:** El texto mediante el cual se genera un enlace entrante a la página analizada. Identificado entre las etiquetas <a> y </a>. Para la presente investigación se utiliza este valor representado por el porcentaje en el cual aparece la keyword seleccionada dentro de la muestra total de enlaces entrantes.
3. **PageRank:** fue planteado como un modelo del comportamiento del usuario, [1] se planteó como **la probabilidad de que un usuario visite una determinada Pagina Web**, es importante mencionar, que una página puede tener un alto PR si hay muchas páginas que la enlazan, o si hay algunas páginas que

la enlazan, y estas tienen un alto PageRank. La forma de calcularlo es la siguiente:

- Se asume que la página A tiene páginas T1..Tn que la apuntan.
- El parámetro d es un factor que puede ser 0 o 1. (Generalmente 0.85)
- C(A) es definido como el número de links salientes de A.

El PageRank de una página A se calcula de la siguiente forma:

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

4. **Edad de una Web:** No hace referencia a la fecha de registraci3n del dominio, sino el lanzamiento de contenido indexable por los buscadores (esto puede cambiar si el dominio cambia de due1o). La forma presentada de medirlo es dd-mm-aaaa.
5. **Cantidad de Dominios que enlazan la Web.** Se refiere a la cantidad de enlaces (medidos en cantidad de dominios) que enlazan a la página objetivo.
6. **Densidad de la keyword en el texto de la página.** La cantidad de veces que se utiliza la keyword seleccionada en el texto visible de la página, medido en porcentaje (%).
7. **Páginas indexadas en Google de ese dominio:** cantidad de páginas de cada dominio que Google tiene indexadas dentro de su índice. El número entero que surge de ejecutar el comando site: dominio.com, reemplazando dominio.com por el dominio requerido.
8. **Uso de los meta keyword y description:** las dos etiquetas que se colocan en el encabezado de una página HTML que tienen como finalidad “dialogar” con Google para informar al buscador acerca de la descripci3n de la página y de sus principales keywords.

La metodología que se utilizó para realizar la experimentaci3n es la siguiente:

1. Entrevistar 30 personas para que piensen la primera palabra que se les ocurra en el idioma inglés.
2. Definir una palabra clave (keyword) genérica en base a la mayor cantidad de repeticiones de una misma palabra. (no términos de búsqueda complejos, solo una keyword que supone un término común entre los encuestados.).
3. Tomar una ventana de “n” casos extraídos de los resultados para esa keyword en un momento dado.

- Analizar esos resultados y compararlos teniendo en cuenta cada uno de los factores que forman la hipótesis.
- Tomar varias muestras durante un periodo de “N” días y comparar los resultados.

#### IV. ENCUESTA PARA DETERMINAR KEYWORD

Se encuestaron 30 personas hispanos parlantes para determinar cual sería la keyword a utilizar, cada una de las personas eligió una palabra en el idioma inglés. Los resultados de la encuesta se muestran en la tabla 1.

Palabra	Repeticiones
House	5
<b>Computer</b>	<b>6</b>
Friend	2
marketing	2
sky	1
car	5
desk	2
women	1
street	1
doll	1
man	1
calculator	1
chief	1
master	1

Tabla 1. Resultados de la encuesta inicial

La palabra **Computer** la mencionaron 6 de las 30 personas encuestadas (un 20%), seguida por House con 5 (un 16.6%), y Car con 5 (un 16.6%), con lo cual ha sido seleccionada para utilizar como keyword para el experimento.

#### V. CAPTURA Y PROCESAMIENTO DE DATOS

El proceso de captura y análisis de datos se realizó de acuerdo a la metodología planteada.

- Se seleccionó la keyword “**Computer**” en base a los resultados de la encuesta realizada.
- Se tomó una ventana de 30 resultados de Google para esa keyword en un momento dado.
- Se analizaron y compararon los resultados teniendo en cuenta cada uno de los factores que forman la hipótesis.

- Se tomaron 3 muestras durante un periodo de 10 días y al comparar los resultados se identificó que la muestra no había variado.

En base a los resultados del primer período de tiempo, se decidió tomar una sola muestra y realizar un análisis estadístico sobre la misma, ya que no existían cambios dentro del período determinado.

Para la obtención de los datos se analizaron diversas herramientas, cada una con sus ventajas y desventajas evidentes. Las herramientas utilizadas varían según el factor a determinar y se detallan en la tabla 2.

Factor (nro)	Herramienta
1	Revisión manual de código HTML
2	SEO Elite
3	Google Bar
4	SEO Quake
5	SEO Elite
6	SEO Quake
7	Google site: dominio.com
8	Revisión manual de código HTML

Tabla 2. Herramientas utilizadas para medir cada factor.

Como se puede observar en la tabla 2, no todos los factores se midieron con la misma herramienta, debido a las limitaciones de cada una y a que fueron desarrolladas para fines específicos, se han utilizado una variedad de las mismas.

#### VI. RESULTADOS

##### A. Evaluación de los resultados obtenidos

Se han obtenido datos para cada uno de los factores en los 30 registros resultantes de la recolección y procesamiento de datos. En principio se realizó una **Correlación de Pearson** para determinar el grado de co-variación entre dos columnas de resultados. Los resultados fueron los siguientes:

Para el **PageRank** (factor 3) con: la cantidad de Enlaces (factor 5), la edad del dominio (factor 4), el anchor text (factor 2), y la cantidad de páginas indexadas en Google (factor 7), **no existe correlación**. Los resultados fueron 0.135148081, -0.247761651, -0.176878223 y 0.135148081 respectivamente.

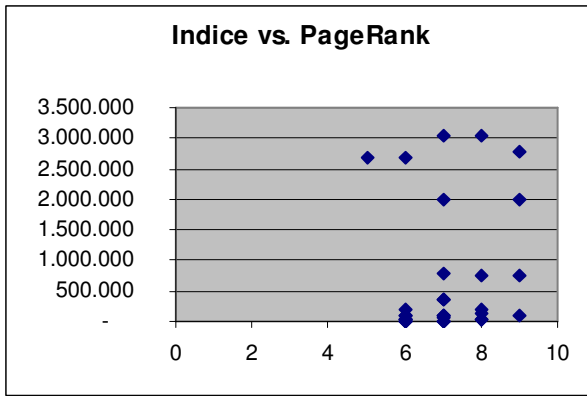


Figura 1. Resultado de la Correlación de Pearson entre el PageRank y la cantidad de páginas indexadas en Google. Como muestra la Figura 1, no se puede establecer esa correlación.

Con respecto al **posicionamiento en Google**, el número de resultado con: la cantidad de Enlaces (factor 5), la edad del dominio (factor 4), el anchor text (factor 2), y la cantidad de páginas indexadas en Google (factor 7), se determina que **sobre la muestra realizada tampoco existe correlación**. Los resultados fueron -0,301004265, -0.107729148, 0.18560831, -0.340572854 respectivamente.

Luego de este análisis se midió la **covarianza** para determinar si los factores tienen relación aunque no sea linealmente, los resultados fueron los siguientes:

Para el **posicionamiento en Google**, o para el **PageRank** el número de resultado, y/o el PR con: la cantidad de Enlaces (factor 5), el anchor text (factor 2), y la cantidad de páginas indexadas en Google (factor 7), **no se puede establecer una variación estadística**.

Comparando estos factores, es decir, el orden de aparición del resultado en Google, y su PageRank, tampoco se puede establecer una relación que indique que cuando un resultado tiene un alto pagerank estará mejor posicionado, no obstante, para esta muestra de 30 resultados, sobre el universo de 689.000.000 resultados para la keyword "Computer", todos los resultados tienen un PageRank igual o mayor que 5.

**Posicion en Google vs. PageRank**

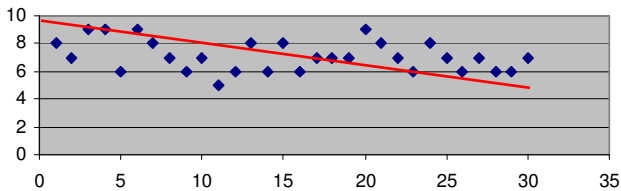


Figura 2. Los puntos marcan la relación entre la posición en los resultados de Google, y el PageRank, la línea roja muestra como debería ser el grafico, si existiera una relación lineal entre estos dos.

Para los factores que tienen resultados binarios, como el primero, y el octavo, no se puede determinar tampoco una relación lineal, en comparación con el orden de aparición en los resultados del buscador.

**B. Análisis de Clusters**

Luego del primer análisis estadístico, se procedió a separar la muestra en 3 clusters, emulando las primeras 3 páginas de resultados de Google, de la siguiente forma:

- Cluster A – Resultados de la posición 1 a la 10.
- Cluster B – Resultados de la posición 11 a la 20.
- Cluster C – Resultados de la posición 21 a la 30.

Se realizó una media de cada uno de los factores en cada cluster, encontrando los siguientes resultados:

**PageRank – ( Factor 3)**

En todos los casos la media del cluster A (7.6) es mayor a la del cluster B (6.9), y la media del Cluster B es mayor a la del cluster C (6.8). Esto se puede observar en la Tabla 3.

Resultados en Google Clusters	Media del PageRank
Cluster A - 1 a 10 resultados	7,6
Cluster B - 11 a 20 resultados	6,9
Cluster C - 21 a 30 resultados	6,8

Tabla 3. La media del PageRank para cada Cluster

**Cantidad de dominios que enlazan la Web (Factor 5)**

En todos los casos la media del cluster A (725.4) es mayor a la del cluster B (640.2), y la media del Cluster B es mayor a la del cluster C (517.0).

Resultados en Google Clusters	Media Factor 5
Cluster A - 1 a 10 resultados	725,4
Cluster B - 11 a 20 resultados	640,2
Cluster C - 21 a 30 resultados	517,0

Tabla 4. La media de la cantidad de dominios que enlazan a la Web para cada Cluster.

**Densidad de la keyword en el texto de la página (Factor 6)**

En todos los casos la media del cluster A (1.29%) es mayor a la media del cluster B (0.62%), y la media del Cluster B es mayor a la del cluster C (0.38%).

Resultados en Google Clusters	Media Factor 6
Cluster A - 1 a 10 resultados	1,29%
Cluster B - 11 a 20 resultados	0,62%
Cluster C - 21 a 30 resultados	0,38%

Tabla 5. La media del porcentaje de repeticiones de la keyword en el texto de la página para cada Cluster.

#### **Páginas indexadas en Google de ese dominio (Factor 7)**

En todos los casos la media en el cluster A es mayor a la media del cluster B, y la media del Cluster B es mayor a la del cluster C.

Resultados en Google Clusters	Media Factor 7
Cluster A - 1 a 10 resultados	1.071.946
Cluster B - 11 a 20 resultados	769.090
Cluster C - 21 a 30 resultados	411.995

Tabla 6. La media de la cantidad de páginas indexadas en Google para cada Cluster.

## VII. CONCLUSIONES

En base al análisis realizado durante la experimentación, luego de la recolección y procesamiento de datos, y con el objetivo de corroborar o refutar los factores para el posicionamiento en Google, se ha detectado que:

**No se comprueba estadísticamente ninguna correlación lineal** entre el orden de un resultado de búsqueda en Google, y los factores planteados.

Tampoco se comprueba que **el solo uso de la keyword en el título, o la colocación de meta tags (factor 8) puedan influir en el posicionamiento.**

Se observa que **en promedio, los primeros 10 resultados de búsqueda poseen un PageRank más alto** que los resultados posicionados entre el 11 y el 20, y entre el 21 y el 30.

También se corrobora **que la densidad de keyword en el texto de la página es mayor para los primeros 10 resultados de búsqueda** que para los resultados posicionados entre el 11 y el 20, y para los resultados que aparecen entre la posición 21 y 30.

El mismo comportamiento se verifica para la cantidad de páginas indexadas en Google por dominio, y la cantidad de enlaces provenientes de otros dominios.

En conclusión, en base a la muestra tomada para este fin, se puede inferir que es posible que dado cada uno de estos tres subgrupos, (los 10 resultados de cada una de las 3 primeras páginas de resultados) un sitio pueda posicionarse entre los primeros subgrupos en Google, si tiene mayor PageRank, más enlaces entrantes, mayor densidad de la keyword en el texto, y mayor número de páginas indexadas en Google, que los otros resultados para esa búsqueda en base a la keyword seleccionada.

Se ha demostrado estadísticamente que estos mismos factores no son fundamentales para posicionar una Web en un subgrupo dado.

Asimismo es importante mencionar, que el objetivo del presente paper es contribuir al conocimiento colectivo detectando casos que pueden re-ponderar la asunción de cuales factores influyen en el posicionamiento orgánico, así como también es posible que también contribuya como material científico al desarrollo de futuras investigaciones.

## VIII. TRABAJO A FUTURO

Sería interesante ampliar este trabajo extendiendo su alcance a varias keywords distintas, en varios idiomas y tomando una muestra de más de 30 resultados por cada una.

Por otra parte también sería un trabajo a futuro ampliar el experimento y la investigación a otros buscadores modernos, no solo utilizar Google como único buscador.

Por último, para poder hacer un análisis completo de la influencia de cada factor en el posicionamiento orgánico en un buscador, es necesario contar con una muestra más grande, basarse en más factores, y realizar cálculos estadísticos complejos para determinar otro tipo de relaciones compuestas tomando varios factores a la vez.

## IX. AGRADECIMIENTOS

Esta investigación no se hubiese podido llevar a cabo sin la participación de Daniela Lopez deluise, quien realizó la tutoría y supervisión.

También agradezco a Sabrina, mi esposa, quien me ayudo en la etapa de experimentación, y análisis de datos.

Asimismo quiero agradecer a todos los bloggers que a través de sus comentarios en mi blog me ayudaron a encontrar el camino, despejando inquietudes en base a su experiencia previa.

## X. REFERENCIAS

### *Papers:*

- [1] Sergey Brin, and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", Stanford University, Stanford, USA, 1998
- [2] Dromey Seamus, Heavin Ciara and Neville Karen, "Combining website search engine optimization with advanced web log analysis", National University of Ireland - Cork, O'Rahilly Building, UCC, Ireland.
- [3] Junghoo Cho, Hector Garcia-Molina "The Evolution of the Web and Implications for an Incremental Crawler", Department of Computer Science, Stanford, Stanford, USA.
- [4] Sepandar D. Kamvar, Taher H. Haveliwala, Christopher D. Manning, Gene H. Golub, "Extrapolation Methods for Accelerating PageRank Computations", Stanford University, Stanford, USA.
- [5] Judit Bar-Ilan, Mark Levene and Mazlita Mat-Hassan, "Dynamics of Search Engine Rankings – A Case Study", The Hebrew University of Jerusalem and Bar-Ilan University, Israel, and the School of Computer Science and Information Systems Birkbeck, University of London
- [6] arvinda, cho, hector, paepcke, rsram, "Searching the Web", Computer Science Department, Stanford University, USA, 2001

- [7] Amy N. Langville and Carl D. Meyer, “Fiddling with PageRank”, Department of Mathematics, N. Carolina State University, Raleigh, USA, August 15, 2003
- [8] Allan Borodin, Gareth O. Roberts, Jeffrey S. Rosenthal, Panayiotis Tsaparas, “Link Analysis Ranking Algorithms, Theory, and Experiments”, 2004
- [9] Junghoo Cho, Sourashis Roy, “Impact of Search Engines on Page Popularity”, UCLA Computer Science Department, USA, 2004
- [10] Mehran Sahami, Vibhu Mittal, Shumeet Baluja, Henry Rowley, “The Happy Searcher: Challenges in Web Information Retrieval” Google Inc., USA, 2004.
- [11] Baoning Wu and Brian D. Davison, “Cloaking and Redirection: A Preliminary Study”, Computer Science & Engineering, Lehigh University, USA, 2005.
- [12] Albert Bifet, Carlos Castillo, Paul-Alexandru Chirita, Ingmar Weber, “An Analysis of Factors Used in Search Engine Ranking”, Technical University of Catalonia, University of Chile, España, Chile, Dinamarca, 2005.
- [13] Michael Brinkmeier, “PAGERANK REVISITED”, Technical University of Ilmenau, Institute for Theoretical and Technical Computer Science, 2006

*Books:*

- [14] John Battelle, “The Search”, Portfolio, 2005.
- [15] Rael Dornfest, Paul Bausch and Tara Calishain, “Google Hacks”, O’Reilly, August 2006.
- [16] Mike Moran and Bill Hunt, “Search Engine Marketing inc.”, IBM Press, August 2006.

*Web Sites:*

- [17] <http://www.seomoz.org/article/search-ranking-factors>