

Factores de Valoración del Texto de los Enlaces I

Una aproximación a la ordenación de documentos basada únicamente en factores que rodean el texto de los enlaces que los apuntan.

Introducción:

En el posicionamiento orgánico actual los enlaces juegan un papel decisivo. Los algoritmos de los motores de búsqueda conceden una importancia vital a los enlaces que apuntan hacia las páginas web a la hora de decidir la relevancia y el orden de los resultados para cualquier consulta.

El valor de una página web se establece no sólo en torno a sus contenidos, sino también a los enlaces que nos apuntan desde otros dominios. Los factores más conocidos en torno a los enlaces son tres:

- La cantidad de enlaces que apuntan hacia una página web.
- La calidad de los enlaces. Cuanto mayor sea el número de enlaces de páginas cercanas en cuanto a la temática de un sitio, más importante será la web apuntada para ese sector.
- El texto de los enlaces: un factor decisivo. El texto de los enlaces que nos apuntan, conocidos por todos como *anchor text*, tienen un peso especial en los algoritmos de los motores de búsqueda actuales.

Los tres puntos tienen mucho que comentar: la cantidad de enlaces tiene un peso variable, normalmente en función de procedencias diversas (Ips, dominios, ubicaciones geográficas, idiomas, etc); la calidad varía en función de los títulos de las páginas que nos apuntan, de lo bien posicionadas que estén esas páginas para las búsquedas del sector, de las antigüedades de los dominios, entre otros factores.

Nos vamos a olvidar de todo esto y nos centraremos únicamente en el análisis del texto del enlace como factor.

La Lista con los Textos de los Enlaces.

Dado un documento arbitrario cualquiera (típicamente, una página principal de un dominio) encontraremos que tiene un determinado número de enlaces desde otros documentos. Si escogemos los enlaces que son válidos para buscadores, cada uno de ellos tendrá asociado un *anchor text*. Imaginemos que hacemos una lista con estos textos de los enlaces.

Ejemplo:

Anchor Text 1: *ordenadores*

Anchor Text 2: *ordenadores portátiles*

Anchor Text 3: *ordenador portátil*

La lista total con las palabras de los textos de los enlaces sería:

Ordenadores, ordenadores, portátiles, ordenador, portátil.

Las Repeticiones.

Si juntamos todas las palabras de forma individual, y a las que se repiten les asociamos un índice con el número de repeticiones que se dan a lo largo de toda la lista, tendremos un conjunto de palabras únicas y con un indicador del número de repeticiones de cada una.

Cuanto mayor sea el índice de una palabra, mayor relevancia tendrá el documento para las búsquedas que la incluyan.

Ejemplo:

Anchor Text 1: *Coches*

Anchor Text 2: *Motos y Coches*

Anchor Text 3: *Motores y Vehículos*

Anchor Text 4: *Motos*

La lista de palabras que tendríamos sería:

Coches, motos, y, coches, motores, y, vehículos, motos.

Ahora eliminamos las repeticiones asociando un índice a cada palabra que determina las repeticiones.

Coches (2)

Motos (2)

Y (2)

Motores (1)

Vehículos (1)

Las Densidades Globales.

Si hacemos una lista de todas las palabras que contienen los textos de los enlaces contando también las repeticiones tendremos un número total de palabras que vamos a denominar n . Para una palabra concreta tendremos un número total de repeticiones que vamos a denotar por i . La densidad de una palabra concreta va a ser el cociente entre i y n .

El porcentaje de densidad lo representamos por $i / n * 100$.

Cuanto mayor sea la densidad de una palabra mayor podría ser la relevancia para las búsquedas que la contengan.

Siguiendo con el ejemplo anterior el número de palabras totales en los textos de los enlaces es de 8. Asociamos a cada palabra su densidad:

Coches $2 / 8 * 100 = 25\%$

Motos $2 / 8 * 100 = 25\%$

Y $2 / 8 * 100 = 25\%$

Motores $1 / 8 = 12.5\%$

Vehículos $1 / 8 = 12.5\%$

Estas densidades no son un factor a tener en cuenta de forma directa para medir la relevancia de un documento en una búsqueda. Pero sí pueden ser útiles para determinar *spam* y estructuras de enlaces artificiales. Las densidades que superaran el 75% para un mínimo de enlaces son completamente anómalas. Las páginas que ocupan primeras posiciones en las búsquedas hispanas más competidas suelen tener su palabra clave principal repetida en torno al 50% de las veces en el anchor.

Repeticiones Ponderadas.

El texto de los enlaces puede tener una o más palabras. Cuando el texto de un enlace está compuesto de una sola palabra debería tener más peso que si esa palabra está incluida en un texto del enlace con varias palabras. El cálculo para este caso sería dividir el número de repeticiones de la palabra entre el número de palabras totales que componen ese anchor.

Ejemplo:

Anchor Text 1: *Hoteles*

Anchor Text 2: *Hoteles en Ciudades*

Anchor Text 3: *Hoteles en Capitales de Provincia*

En el caso 1 *Hoteles* tendría un peso del 100% y contaría con un peso de una unidad a efectos de repeticiones.

En el caso 2 *Hoteles* tendría un peso del 33% y su peso sería de $1 / 3$.

En el caso 3 el peso de la palabra *Hoteles* sería de 1 repetición / 5 palabras: 0.20 unidades.

Si ahora aplicamos el número de repeticiones asociando a cada palabra su peso, llegamos al cálculo de las repeticiones ponderadas.

Volviendo al ejemplo:

Hoteles: $1 + 1 / 3 + 1 / 5$

En: $1 / 3 + 1 / 5$

Ciudades: $1 / 3$

Capitales: $1 / 5$

De: $1 / 5$

Provincia: $1 / 5$

La puntuación obtenida para cada palabra clave reflejaría la relevancia de la misma para las consultas que conllevan esa palabra exacta y es un método de ordenación con buenos resultados. Su utilización debería ser utilizada en consonancia con las densidades globales.

Continuará...