

UNIVERSIDAD DE OVIEDO



ESCUELA DE INGENIERÍA INFORMÁTICA

TRABAJO FIN DE MÁSTER

“Métrica de factores on-page en el posicionamiento de páginas web en los motores de búsqueda orgánicos”

DIRECTORES:

**B. CRISTINA PELAYO GARCÍA-BUSTELO
EDWARD ROLANDO NÚÑEZ-VALDÉZ**

AUTOR: MANUEL BÉCARES PÉREZ

Vº Bº del Director del
Proyecto

Resumen

Actualmente Internet se ha convertido en el medio más utilizado por personas, instituciones y empresas para darse a conocer. Además, generalmente, las personas acuden a la red de redes para buscar prácticamente cualquier tipo de información.

Debido al crecimiento de la web, cada vez toman mayor importancia los motores de búsqueda en la red, los cuales gracias a sus optimizados algoritmos son capaces de clasificar la información y ordenarla por relevancia.

Generalmente, cuando se quiere información sobre un tema se busca información en Internet a través de un buscador. Por ello, cualquier persona, institución o empresa que quiere darse a conocer en Internet debe garantizar su presencia en los principales buscadores.

Esta presencia no se consigue sin tener un buen producto o información que ofrecer pero, sobre todo, sin un trabajo de optimización del posicionamiento en los buscadores (*SEO - Search Engine Optimization*) que ayude a que esta aparezca en las primeras posiciones de los resultados de las búsquedas.

En relación al SEO, los buscadores no aportan casi ninguna información y, de hecho, es uno de sus secretos mejor guardados. Además en el mundo científico, debido a este secretismo, existen pocas fuentes de información e investigaciones rigurosas.

El objetivo de este trabajo de investigación es estudiar los factores empleados por los buscadores para ordenar los resultados de una búsqueda, y definir una métrica que permita optimizar el posicionamiento de las páginas web en dichos resultados.

Palabras Clave

Buscadores, SEO, SERPs, Factores *on-page*, Búsqueda orgánica, Posicionamiento, Página web.

Abstract

Nowadays, Internet has become the most used way by people, institutions and companies to get popularity. People use the Internet to find any type of information.

Due to this growth search engines are also growing their importance on the Internet, and thanks to its optimized algorithms they now are able to sort all the information correctly by relevance.

When people want information on a topic, they search for information on the Internet through a search engine. Therefore any person, institution or company that wants to get noticed on the Internet should increment its presence in major search engines.

This presence on search engines can't be achieved without having a good product or information to offer. However, to get noticed in the major search engines is very difficult without Search Engine Optimization (*SEO*).

Search engines don't provide much information about *SEO* and, in fact, this kind of information is top secret. In the scientific world, because of this secrecy, there are few sources of information and rigorous research about *SEO*.

The objective of this research is to discover some of the best kept secrets by search engines in relation to *SEO* tools and practices. There is very little information about this field, due to the secrecy that surrounds this topic.

Keywords

Search engines, SEO, SERPs, *On-page* factors, Organic search, Ranking, Web page.

Índice General

CAPÍTULO 1. INTRODUCCIÓN	17
1.1 PLANTEAMIENTO Y JUSTIFICACIÓN DEL TRABAJO	18
1.1.1 <i>El posicionamiento en los buscadores</i>	18
1.2 HIPÓTESIS Y OBJETIVOS DE LA INVESTIGACIÓN	22
1.2.1 <i>Objetivos</i>	23
1.3 METODOLOGÍA A SEGUIR DURANTE LA INVESTIGACIÓN	25
CAPÍTULO 2. ESTADO DEL ARTE	29
2.1 DEFINICIÓN DE SEO	30
2.2 CONCEPTO DE CONVERSIÓN	31
2.3 CICLO DE VIDA DEL <i>SEO</i>	32
2.4 FACTORES <i>ON-PAGE</i>	34
2.4.1 <i>Dominio</i>	34
2.4.2 <i>Servidor</i>	34
2.4.3 <i>Arquitectura</i>	34
2.4.4 <i>Contenido</i>	35
2.4.5 <i>Enlaces internos</i>	35
2.4.6 <i>Elementos del sitio web</i>	35
2.4.7 <i>Factores específicos de cada pantalla</i>	35
2.4.8 <i>Uso de palabras clave y su prominencia</i>	36
2.4.9 <i>Enlaces salientes</i>	36
2.5 FACTORES <i>OFF-PAGE</i>	37
2.5.1 <i>Enlaces entrantes</i>	37
2.5.2 <i>Enlaces entrantes por separado</i>	37
2.5.3 <i>Perfil de los Visitantes y conducta</i>	38
2.6 BUENAS PRÁCTICAS DE <i>SEO</i> SEGÚN GOOGLE.....	38
2.7 TÉCNICAS <i>SEO</i> DE SOMBRERO BLANCO Y TÉCNICAS <i>SEO</i> DE SOMBRERO NEGRO.....	40
2.8 INGENIERÍA INVERSA Y <i>SEO</i>	41
2.9 IDEAS ADOPTADAS DEL ESTADO DEL ARTE.....	41
2.9.1 <i>Artículo “Cómo mejorar tu ranking en Google: Mitos y realidades”</i>	42
2.9.2 <i>Matices de la investigación</i>	42
CAPÍTULO 3. DESARROLLO DE LA INVESTIGACIÓN.....	47
3.1 FACTORES A ESTUDIAR EN LA INVESTIGACIÓN.....	48
3.2 TÉCNICA PARA AISLAR LOS FACTORES A ESTUDIAR	53
3.3 MOTORES DE BÚSQUEDA SELECCIONADOS	55
3.4 RIESGOS DEL PROYECTO DE INVESTIGACIÓN	55
3.4.1 <i>Que los buscadores no indexen las muestras</i>	55
3.4.2 <i>Que los buscadores indexen tarde las páginas web de las muestras</i>	57
3.4.3 <i>Que los resultados no revelen ningún patrón</i>	57
3.5 PROTOTIPO DEL GENERADOR DE PÁGINAS WEB	58
3.5.1 <i>Arquitectura del prototipo</i>	59
3.5.2 <i>Valores de entrada del programa</i>	63
3.5.3 <i>Formato y organización de los textos de muestra</i>	63
3.5.4 <i>Aleatoriedad en las páginas web generadas</i>	65
3.5.5 <i>Resultados del prototipo</i>	71

3.6	ELECCIÓN Y CREACIÓN DE MUESTRAS DE PALABRAS CLAVE A ESTUDIAR	74
3.6.1	<i>Palabras clave elegidas</i>	74
3.6.2	<i>Creación de los ficheros de muestra</i>	74
3.6.3	<i>Ejemplo de ficheros de muestra</i>	75
3.7	INDEXACIÓN DE LAS MUESTRAS CREADAS	87
3.7.1	<i>Publicación de páginas web en Internet</i>	88
3.8	CÁLCULO DE LA MÉTRICA	89
3.9	OBTENCIÓN DE RESULTADOS	91
3.9.1	<i>Palabra clave “dieta”</i>	92
3.9.2	<i>Palabra clave “adelgazar”</i>	94
3.9.3	<i>Palabra clave “computador”</i>	96
3.9.4	<i>Palabra clave “car”</i>	98
3.10	CÁLCULO DE LA MÉTRICA	100
3.10.1	<i>Primera iteración</i>	100
3.10.2	<i>Segunda iteración</i>	102
3.10.3	<i>Tercera iteración</i>	103
3.10.4	<i>Tabla final de resultados</i>	104
3.11	GRÁFICAS DE RESULTADOS	107
3.11.1	<i>Gráfica de puntuaciones obtenidas para la palabra clave “dieta”</i>	107
3.11.2	<i>Gráfica de puntuaciones obtenidas para la palabra clave “adelgazar”</i>	108
3.11.3	<i>Gráfica de puntuaciones obtenidas para la palabra clave “computador”</i>	109
3.11.4	<i>Gráfica de puntuaciones obtenidas para la palabra clave “car”</i>	109
CAPÍTULO 4. CONCLUSIONES Y TRABAJO FUTURO		113
4.1	HIPÓTESIS.....	113
4.2	OBJETIVOS	113
4.2.1	<i>Objetivo principal</i>	113
4.2.2	<i>Generación de muestras de páginas web</i>	114
4.2.3	<i>Indexación de páginas web</i>	114
4.2.4	<i>Obtención de la métrica que siguen los buscadores</i>	114
4.3	CONCLUSIONES ACERCA DE LOS FACTORES ESTUDIADOS	115
4.4	RECOMENDACIONES PARA MEJORAR EL POSICIONAMIENTO DE LAS PÁGINAS WEB	116
4.5	DIFUSIÓN DE LOS RESULTADOS.....	117
4.6	TRABAJO FUTURO.....	117
CAPÍTULO 5. GLOSARIO.....		121
CAPÍTULO 6. REFERENCIAS BIBLIOGRÁFICAS.....		123
CAPÍTULO 7. APÉNDICES		125
7.1	CONTENIDO ENTREGADO EN EL CD-ROM	125
7.1.1	<i>Prototipo</i>	125
7.1.2	<i>Artículo</i>	126
7.2	ÍNDICE ALFABÉTICO	127

Índice de Figuras

Ilustración 1 – Cantidad de usuarios en Internet en el año 2011	17
Ilustración 2 – Porcentaje de clics en las SERP de los motores de búsqueda por posición.	18
Ilustración 3 - Search Engine Result Page (SERP) Así es una página de resultados de un buscador	19
Ilustración 4 – Mapa de calor de una SERP en Google, el principal buscador web actualmente(2009). 20	
Ilustración 5 – Ciclo de vida del SEO	32
Ilustración 6 – Resultados para la palabra clave “Smartphone” en http://google.es	49
Ilustración 7 – Resultados para la palabra clave “Smartphone” en http://google.co.uk	49
Ilustración 8 – Resultados de búsqueda de la palabra clave “Tablet” en Google	50
Ilustración 9 – Resultados de búsqueda de la palabra clave “Tablet” en Yahoo omitiendo los resultados publicitarios	50
Ilustración 10 – Resultados de búsqueda de la palabra clave “Tablet” en Bing omitiendo los resultados publicitarios.....	51
Ilustración 11 – Ejemplo de página de índice para una muestra de páginas web de la palabra clave “informática”	54
Ilustración 12 – Estadísticas de uso de motores de búsqueda en PC durante el año 2012	55
Ilustración 13 – Cómo saber los enlaces que apuntan a http://www.uniovi.es	56
Diagrama 3.1 – Diagrama de paquetes del prototipo	59
Diagrama 3.2 – Diagrama de clases del modelo del prototipo	60
Diagrama 3.3 – Diagrama de la capa de negocio del prototipo.....	61
Diagrama 3.4 – Diagrama de la capa de persistencia del prototipo	62
Ilustración 14 – Cómo son las muestras de texto de entrada optimizadas.....	64
Ilustración 15 – Cómo son las muestras de texto de entrada sin optimizar	64
Ilustración 16 – Ficheros de entrada para cada palabra clave.....	65
Ilustración 17 – Páginas web generadas por el prototipo para la palabra clave “noticias”	71
Ilustración 18 – Apariencia de una página web generada no optimizada	72
Ilustración 19 – Apariencia de una página web generada optimizada	72
Ilustración 20 –Archivos necesarios para generar muestras de cada palabra clave	75
Ilustración 21 – Ejecución del prototipo desde Eclipse	83
Ilustración 22 – Páginas generadas para la palabra clave “Computador”	84
Ilustración 23 – Metadatos generados para una página web de la palabra clave “Computador”	86
Ilustración 24 – Formulario de envío de páginas web a Google	87
Ilustración 25 – Formulario de envío de URL a Bing.....	88
Tabla 1 – Plantilla para rellenar puntuaciones de los factores a estudiar para cada página web de una búsqueda.....	90
Ilustración 26 – 5 primeros resultados en Google para la palabra clave “dieta” de nuestro conjunto de muestras.....	92
Tabla 2 – Tabla de resultados para la palabra clave “dieta”	93
Ilustración 27 - 5 primeros resultados en Google para la palabra clave “adelgazar” de nuestro conjunto de muestras.....	94
Tabla 3 - Tabla de resultados para la palabra clave “adelgazar”	95
Ilustración 28 - 5 primeros resultados en Google para la palabra clave “computador” de nuestro conjunto de muestras	96
Tabla 4 - Tabla de resultados para la palabra clave “computador”	97
Ilustración 29 - 5 primeros resultados en Google para la palabra clave “car” de nuestro conjunto de muestras.....	98
Tabla 5 - Tabla de resultados para la palabra clave “car”	99
Tabla 6 – Primera estimación de puntuación de las páginas web de cada búsqueda.....	102

Tabla 7 – Resumen de la primera estimación de las puntuaciones	102
Tabla 8 – Resumen de la segunda estimación de las puntuaciones.....	103
Tabla 9 – Tabla final de puntuaciones con los pesos finales asignados a cada factor	105
Tabla 10 – Pesos asignados a cada factor de la métrica	106
Gráfica 1 – Puntuaciones y posiciones para la palabra clave “dieta”	107
Gráfica 2 - Puntuaciones y posiciones para la palabra clave “adelgazar”	108
Gráfica 3 - Puntuaciones y posiciones para la palabra clave “computador”	109
Gráfica 4 - Puntuaciones y posiciones para la palabra clave “car”.....	109
Tabla 11 – Pesos finales de cada factor de la métrica obtenida	115

PARTE I - INTRODUCCIÓN

Capítulo 1. Introducción

Internet se ha convertido en el mayor medio de difusión de información en la actualidad. Miles de millones de usuarios se conectan cada día a la red de redes en busca de información o productos.

Actualmente, según datos oficiales de Internet World Stats, y tal como se puede ver en la Ilustración 1 existen más de dos mil doscientos millones de usuarios que se conectan a Internet en todo el mundo (Internet World Stats 2011), lo cual significa que un tercio de la población mundial alguna vez se ha conectado a Internet.

Con tal cantidad de usuarios y potenciales clientes no son pocas las personas, instituciones y empresas que quieren darse a conocer en Internet, y es por ello que el número de páginas web que hay en la actualidad es abrumador, y el crecimiento de las mismas es exponencial.

Debido al número de páginas web que hay actualmente, cuyas cifras son imposibles de medir con exactitud, pero se calcula que haya más de 550 millones de dominios registrados, en la actualidad toman especial importancia los motores de búsqueda.

Los motores de búsqueda orgánicos son un tipo de software que se encarga de coleccionar datos de páginas web (URL, palabra o conjunto de palabras clave que aparecen en la página, estructura del código e hiperenlaces que apuntan a esa web en Internet). Todos esos datos son indexados y almacenados en bases de datos de gran tamaño. A partir de estas estructuras de datos, los motores de búsqueda orgánicos son capaces de, cuando se realiza una consulta a los mismos, devolver un conjunto de resultados en forma de páginas web ordenadas por relevancia (Yalçın and Köse 2010).

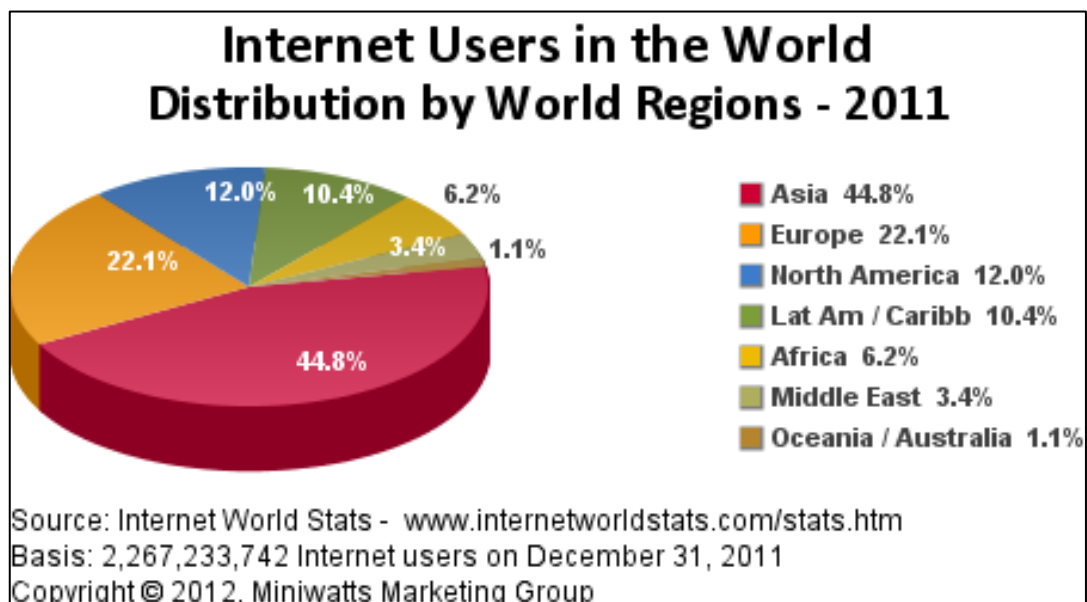


Ilustración 1 – Cantidad de usuarios en Internet en el año 2011

1.1 Planteamiento y justificación del trabajo

La motivación principal de este trabajo de investigación es encontrar las pautas que se deben seguir para garantizar que una determinada página web aparezca en la mejor posición en los buscadores.

1.1.1 El posicionamiento en los buscadores

Con el crecimiento comentado anteriormente de Internet y de las páginas web, y con la cantidad de sitios web que ya existen actualmente, sería imposible imaginarnos encontrar cada cosa que busquemos en Internet sin utilizar los motores de búsqueda.

Aproximadamente el 85% de usuarios, cuando quieren encontrar algo en Internet, acuden a los buscadores para ello (Kobayashi and Takeda 2001).

Es por eso que para darse a conocer en Internet a través de una página web cualquier persona, institución o empresa necesita **ser relevante para los buscadores**, es decir, aparecer en las primeras posiciones para las búsquedas con palabras clave que le interesen que estén relacionadas con la información o productos que ofrecen.

Debido a la enorme competencia existente en un mercado tan grande como es el de Internet, salvo que lo que se ofrezca en una página web sea algo totalmente distinto e innovador que no se parezca a nada de lo que ya había, **aparecer en las primeras posiciones de una búsqueda** por una o varias palabras clave en los buscadores es una tarea sumamente difícil que en la mayoría de los casos requiere de expertos que hagan este trabajo.

Tal y como puede verse en la Ilustración 2, cuando un usuario realiza una búsqueda en un motor de búsqueda, en el 53% de las ocasiones hace clic en el primer resultado, en el 15% de ocasiones en el Segundo y el 9% en el tercero. La cuarta posición se lleva un 6% de clics y tan solo el 5% de clics se dan desde la quinta posición en adelante (Compete 2012).

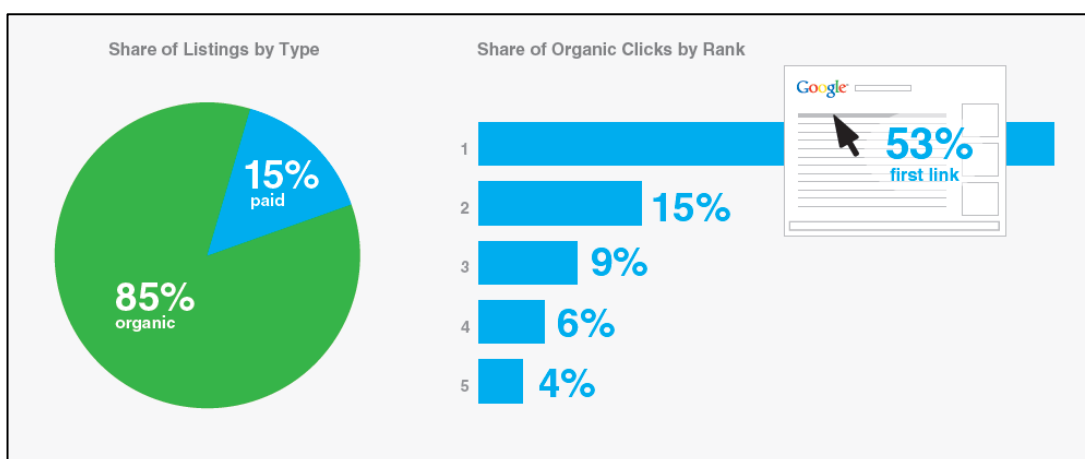


Ilustración 2 – Porcentaje de clics en las SERP de los motores de búsqueda por posición.

En la Ilustración 3 podemos ver la anatomía de una página de resultados de búsqueda, que de ahora en adelante llamaremos *SERP* (Search Engine Result Page).

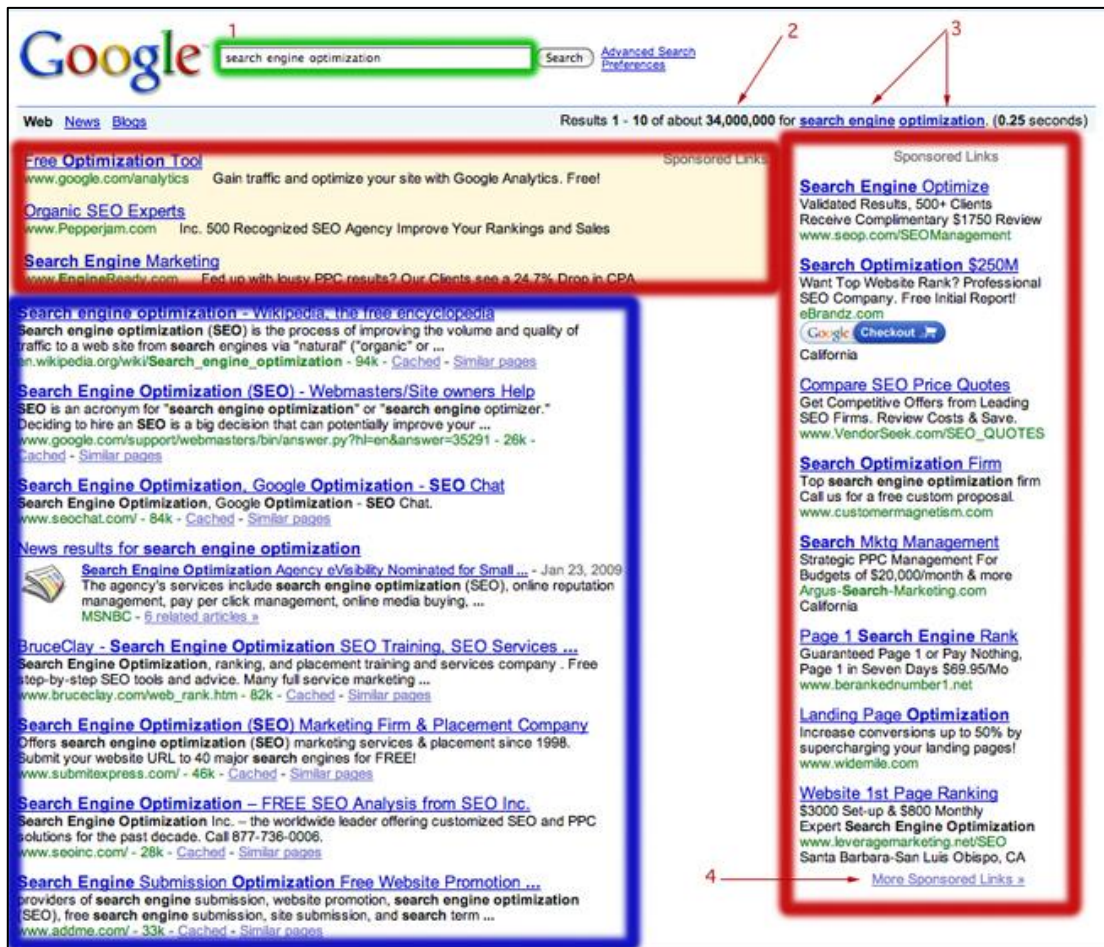


Ilustración 3 - Search Engine Result Page (SERP) Así es una página de resultados de un buscador

Como se puede apreciar, hay tres bloques principales de resultados. El bloque central superior y el bloque lateral son resultados de búsqueda de pago. Estos resultados son excluidos de esta investigación. En el bloque central se ven los **resultados de búsqueda orgánicos**, es decir, los que no son de pago y aparecen ahí de forma natural.

A continuación, en la Ilustración 4, una imagen del mapa de calor de las búsquedas en un motor de búsqueda de páginas web:

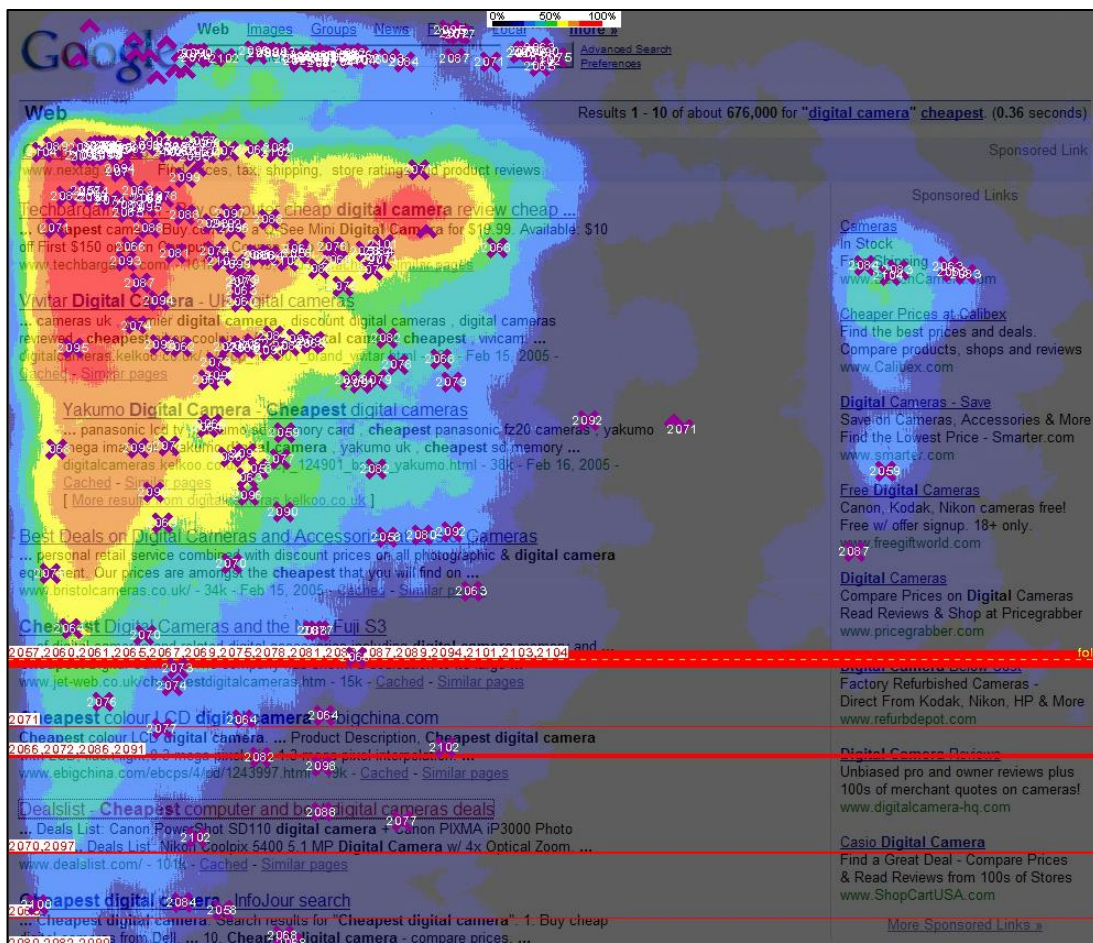


Ilustración 4 – Mapa de calor de una SERP en Google, el principal buscador web actualmente(2009).

Como vemos en la imagen superior, proporcionada por el propio Google, el mapa de calor de las SERP tiene una zona muy clara en la que el usuario centra la vista en el primer momento, que es la zona en la que aparecen los tres primeros resultados de búsqueda (Peñarrolla 2009).

Actualmente las SERP varían mucho su estructura dependiendo del tipo de búsqueda que se realice y de los resultados devueltos, pero en esencia el mapa de calor sigue teniendo su punto más visto por los usuarios alrededor de los tres primeros resultados de búsqueda, sin tener en cuenta los resultados de pago, los cuales son inamovibles.

Con todos estos datos queda claro que para ser visible en Internet hay dos opciones:

1. Pagar para aparecer en los anuncios, para lo cual existe un campo de estudio que se centra en esto exclusivamente (SEM) y que también es muy complejo.
2. Intentar aparecer en los tres primeros resultados de búsqueda orgánicos, donde entra en juego el SEO.

Search Engine Optimization (SEO) es el proceso de identificación de factores de una página web que podrían afectar a la accesibilidad de los motores de búsqueda a la misma y afinar los muchos elementos de dicha página web para que esta pueda alcanzar la visibilidad más alta posible cuando un motor de búsqueda atiende una consulta de palabras clave por las cuales se puede indexar dicha web(Zhang and Dimitroff 2005).

El mundo del *SEO* está envuelto por un halo de secretismo el cual hace que no existan prácticamente estudios científicos a través de los cuales se pueda demostrar que determinados factores afectan al posicionamiento de una página web, y es por eso que surge la idea de realizar este trabajo de investigación.

La idea de esta investigación es intentar obtener una **métrica de algunos factores que emplean los buscadores para ordenar las páginas web por relevancia** cuando se realiza una búsqueda.

Estos factores se dividen en dos tipos(Zhang and Dimitroff 2005):

- Factores internos de la página (*on-page*).
- Factores externos a la página (*off-page*).

En este proyecto de investigación va a limitarse el número de factores a estudiar. Se estudiarán algunos de los factores internos de la página. Dentro de estos factores, podemos distinguir dos sub-tipos(Zhang and Dimitroff 2005):

- Factores de la estructura del código.
- Factores del servidor y del dominio.

El estudio se centrará en los factores de la estructura del código, dado que incluir otros factores haría imposible obtener conclusiones sobre la relevancia de una página web frente a otra, al estar mezclando factores independientes.

La idea de realizar esta investigación además surge debido a que apenas hay avances en este campo de la informática tan importante como es hoy en día con la creciente evolución de Internet, como ya explicamos anteriormente. Pero es que además, las pocas investigaciones realizadas, son muy generalistas y tienen en cuenta infinidad de factores simultáneamente, los cuales influyen unos en otros y no se pueden extraer conclusiones claras acerca del peso de cada uno. Otras investigaciones simplemente realizan afirmaciones gratuitas sin aportar ninguna prueba de ello.

Por ello **en esta investigación se ha optado por hacer un estudio a fondo de los principales factores *on-page* del SEO** en lugar de estudiar todos los factores que influyen en el posicionamiento de una web de forma simultánea.

La principal aportación de esta investigación es conseguir medir algunos de estos factores, descritos en el siguiente apartado, de forma aislada sin que el resto de factores intervengan, y así obtener una métrica realista de los factores *on-page* que emplean los motores de búsqueda.

1.2 Hipótesis y objetivos de la investigación

La hipótesis inicial de la investigación es que existe una métrica que siguen los buscadores a través de la cual son capaces de ordenar los resultados de búsqueda para una determinada palabra clave por relevancia a partir de determinados factores, y que esta métrica puede averiguarse a través de ingeniería inversa.

Los factores a estudiar, como ya se comentó en el planteamiento del proyecto, van a ser únicamente los factores internos que afectan a la estructura de una página web, luego la hipótesis planteada anteriormente va a estar estrechamente relacionada con estos factores (Zhu and Wu 2011),(Google 2012),(Cornello 2010):

1. Que la palabra clave aparezca en el título de la página.
2. El número de apariciones de la palabra clave en el título de la página.
3. Que la palabra clave aparezca en la meta descripción de la página.
4. Que la palabra clave aparezca la meta de título de la página.
5. Que la palabra clave aparezca en la meta de palabras clave de la página.
6. Que la palabra clave aparezca en el encabezado de nivel 1 de la página.
7. Que la palabra clave aparezca en el encabezado de nivel 2 de la página.
8. Que la palabra clave aparezca en el encabezado de nivel 3 de la página.
9. Que la palabra clave aparezca en la primera frase de la página.
10. Que la palabra clave aparezca en la última frase de la página.
11. El número de veces que la palabra clave aparece en negrita/cursiva/subrayada.
12. El número de veces que la palabra clave aparece en enlaces externos.
13. La densidad de la palabra clave en la página.
14. Que la palabra clave aparezca en el texto alternativo de una imagen de la página web.
15. Que la palabra clave aparezca en la URL de la página web.

Estudiando y analizando todos estos factores en diferentes muestras de páginas web deberíamos ser capaces de confirmar o rechazar nuestra hipótesis de que los buscadores utilizan algún tipo de métrica para ordenar los resultados de las páginas web mostradas en las búsquedas.

Hipótesis: Los motores de búsqueda utilizan algún tipo de métrica para ordenar los resultados de búsqueda y es posible aproximarla gracias a la ingeniería inversa.

1.2.1 Objetivos

Teniendo en cuenta las hipótesis, **el objetivo fundamental del proyecto es conseguir averiguar dicha métrica que en teoría siguen los buscadores para ordenar los resultados de búsqueda**, pero para llegar a esto debemos antes preparar el entorno donde van a realizarse las pruebas de los factores mencionados anteriormente, lo cual es quizá la parte más laboriosa de esta investigación:

Objetivo: Obtener una métrica con diferentes factores y pesos para cada factor a través de la cual podamos ordenar por puntuación los resultados de búsqueda.

1.2.1.1 *Objetivo 1: Obtener muestras de páginas web para analizar*

Para poder analizar los diferentes factores de las hipótesis se necesitan muestras de páginas web con ciertas peculiaridades:

- Deberán tener un número razonable de páginas web.
- Deberán estar aisladas de todo tipo de factores externos.
- Cada muestra deberá estar orientada a una única palabra clave.
- Cada página web de la muestra debe ser diferente del resto.
- Debe tenerse información acerca de la estructura de cada página en forma de metadatos.

Teniendo en cuenta estas condiciones que debe cumplir cada muestra, para la investigación no sirve realizar una búsqueda en cualquier buscador sin más y observar los resultados, ya que estos estarían afectados por factores externos.

Es por ello que las muestras deben ser generadas a través de una aplicación que sea capaz de generar miles de páginas web diferentes a través de una palabra clave y diferentes textos, y que además sea capaz de, según si se le dice que genere páginas optimizadas o no, emplear los factores incluidos en las hipótesis para generar las páginas web y guardar en un fichero de metadatos para cada página web sus propiedades para cada factor.

Objetivo: Generación de muestras de páginas web con las condiciones adecuadas para el estudio del posicionamiento de las mismas.

1.2.1.2 *Objetivo 2: Indexar las muestras en diferentes motores de búsqueda*

El segundo objetivo de esta investigación es conseguir que los principales motores de búsqueda existentes indexen las muestras de páginas web. Este objetivo tiene la dificultad de cómo conseguir indexar miles de páginas web en los buscadores, lo cual no es una tarea sencilla, por lo que habrá que encontrar un método que realice esta tarea de forma automática.

Objetivo: Indexación de las diferentes muestras de páginas web generadas en los buscadores.

1.2.1.3 *Objetivo 3: Analizar patrones en las muestras de páginas*

Una vez creadas las muestras de páginas web e indexadas en los principales buscadores, el siguiente objetivo es conseguir identificar patrones que se cumplan en las primeras posiciones devueltas por los buscadores y asociarlos a diferentes factores.

A partir de esto podríamos determinar cuáles son los factores que afectan al posicionamiento de una página web positivamente y cuáles afectan negativamente y, si se dan las condiciones apropiadas, determinar una estimación del peso de cada factor en el algoritmo de ordenación de los buscadores.

Objetivo: Búsqueda de patrones y obtención, a partir de los mismos, de los pesos de cada factor propuesto en busca de la métrica.

1.3 Metodología a seguir durante la investigación

La metodología de investigación que se va a seguir en este trabajo estará muy marcada por los objetivos del mismo. Esta metodología va a ser lineal, objetivo por objetivo siguiendo el orden indicado en el apartado anterior de los objetivos.

Lo primero que se va a investigar es si los factores escogidos son realmente todos los que debemos incluir en nuestra investigación o si hay que añadir o quitar algún factor.

Los factores se han obtenido a partir de diferentes partes del texto de algunas referencias, en las que se mencionan diferentes factores *on-page* que, en teoría, afectan al posicionamiento de las páginas web en los buscadores.

Obtenido el conjunto de factores final, el siguiente paso es construir un prototipo que sea capaz de generar páginas web de forma aleatoria a partir de una palabra clave y unos textos de muestra y que además emplee los factores incluidos en las hipótesis en el algoritmo de cálculo aleatorio de la estructura de la página web.

Realizado dicho prototipo, el siguiente objetivo es indexar en Google, Yahoo y Bing, que son los principales motores de búsqueda actualmente, las páginas web generadas para alguna palabra clave y con unos textos de muestra suficientemente amplios en cuanto a extensión y concretos en cuanto a contenido (la palabra clave elegida en principio no tiene relevancia siempre y cuando tenga relación con los textos que la envuelven).

Una vez indexadas las páginas web generadas en los principales buscadores, el siguiente paso es obtener los resultados realizando una búsqueda que englobe sólo a esas páginas web, y comprobar cómo se han posicionado.

A partir de esa muestra de resultados, hay que deducir cuáles son los factores más importantes para los buscadores en su métrica para ordenar los resultados por relevancia, y verificar si se cumple la hipótesis planteada.

Esta tarea va a realizarse a mano empleando los metadatos que ya se tienen de cada página web generados por el primer prototipo y analizando visualmente diferentes patrones que cumplan las páginas web en las primeras posiciones.

El análisis de estos patrones se realizará de forma numérica, es decir, se crearán tablas indicando qué páginas web cumplen cada factor y, a partir de ahí, intentar asignar pesos que sean capaces de ordenar las páginas por puntuación en función de la posición en la que se encuentren en los resultados de búsqueda.

PARTE II - ESTADO DEL ARTE

Capítulo 2. Estado del arte

El *SEO* es una parte importantísima a tener en cuenta dentro de una estrategia de creación de un sitio web o de llevar un negocio o información a Internet.

El campo de la optimización de páginas web para mejorar su accesibilidad por los motores de búsqueda es un campo que está en continua evolución, principalmente debido a los continuos cambios que introducen los buscadores en sus algoritmos, sobre todo Google que es el motor de búsqueda que reúne la inmensa mayoría del tráfico de Internet.

A pesar de ser un campo en el que en teoría deben existir investigaciones que lo estudien continuamente e intenten descifrar los métodos que emplean los buscadores para indexar y ordenar resultados, no existe mucha información verdaderamente verificada acerca del *SEO*.

2.1 Definición de SEO

Existen diferentes definiciones sobre SEO, y algunas de las más relevantes son:

- *“Conseguir que los buscadores sepan que nuestra web existe y que además consideren que realmente habla de aquello que un usuario dado está buscando y lo coloque lo antes posible entre las miles de respuestas obtenidas es el reto de lo que se ha dado en llamar posicionamiento y optimización en buscadores”*(Larreina 2005).
- *“Podemos definir posicionamiento web como el conjunto de procedimientos y técnicas que tienen como finalidad dotar a un sitio o a una página web de la máxima visibilidad en Internet”*(Codina and Marcos 2005).
- *“SEO es el proceso mediante el cual, siguiendo los principios de los motores de búsqueda en cuanto a la estructura que deben tener los sitios web y las relaciones entre los mismos se consigue una **mejora de la visibilidad de los sitios web en los motores de búsqueda***(Zhao, Lu, and Duan 2009).

En definitiva, “aplicar *SEO*” sobre una página o sitio web es utilizar todos los procedimientos y herramientas que, estando al alcance del encargado de esta labor, contribuyen a conseguir un mejor posicionamiento de dicha página web en los buscadores.

Dentro del campo del *SEO* existen muchos factores que pueden afectar al posicionamiento de páginas web. Según algunos expertos, el algoritmo de Google, que es el que más importancia tiene dentro de Internet, tiene hasta más de 200 factores según muchos expertos(Carlos 2012).

Para el propósito de esta investigación tan solo van a estudiarse algunos de estos factores, pero hay que conocer cuáles son todos ellos, porque habrá que aislar los factores que se van a estudiar del resto.

2.2 Concepto de conversión

Ya sabemos que el *SEO* es optimizar y mejorar el posicionamiento de una página web en los buscadores, pero esto no puede hacerse de cualquier forma.

En algunos lugares con poca autoridad se menciona que el *SEO* es conseguir más visitas para una página web, pero el objetivo del *SEO* no es ese, sino que es conseguir más visitas a las que el contenido de una página web les resulte atractivo y relevante.

De nada sirve si de un día para otro conseguimos aumentar el número de visitas a una página web si estas abandonan la misma sin ver el contenido siquiera debido a que no es una página relevante para los usuarios.

El objetivo del *SEO*, además de conseguir visitantes, es que el contenido o producto que ofrece una página web optimizada sea relevante para estos (Yalçın and Köse 2010). ¿Cómo consigue esto el *SEO*? Vamos a estudiarlo en el apartado del ciclo de vida, pero básicamente lo que hace un experto que optimice páginas web es seleccionar minuciosamente las palabras clave por las que posicionar la misma.

En conclusión, es importante conseguir tráfico para una página web. Obviamente el tráfico es lo que da vida a un sitio web, más actualmente estando de moda la Web 2.0 y la interactividad con el usuario. Sin embargo para conseguir que el usuario interactúe con nuestro sitio web es fundamental que este sea atractivo y relevante para el visitante. Es por esto que el objetivo del *SEO* no es solo conseguir visitas, sino que el objetivo de cualquier estrategia de posicionamiento para una página web es conseguir visitantes que interactúen con la página web y sus contenidos, o en caso de vender un producto o servicio, diseñar una estrategia para que el usuario llegue al mismo a través de una serie de pasos, de los cuales el primero y fundamental es captar al usuario en la página del sitio adecuada con las palabras clave adecuadas.

Esto último es importante porque si conseguimos que un usuario entre directamente a la página de pago de nuestro sitio web por ejemplo, seguramente si antes no ha visto información del producto el usuario de al botón de atrás de inmediato.

Conseguir que un visitante, a través de nuestra estrategia de *SEO* y navegación por la web, compre un producto o acceda a una página web determinada es lo que en el mundo *SEO* se conoce como “conversión” (Search Engine Watch 2013).

Así, el objetivo fundamental para cualquier campaña de *SEO* no es conseguir más visitantes, sino conseguir más conversiones.

2.3 Ciclo de vida del SEO

El ciclo de vida de una campaña de SEO sobre una página web ya creada consta de tres fases (Codina 2007):



Ilustración 5 – Ciclo de vida del SEO

1. Análisis:
 - a. Objetivos:
 - i. Determinación de la(s) palabras clave para las cuales se desea optimizar el sitio.
 - ii. Determinación de posibles *partners* que pueden enlazar nuestro sitio (con enlaces recíprocos, por nuestra parte, en su caso).
 - b. Herramientas:
 - i. Análisis y estudio del sitio por un experto humano y/o análisis con programas específicos (p.e., IBP3, Arelis).
 - ii. Benchmarking por comparación de la mejor competencia; realizado por un experto humano o bien un programa específico (o ambas cosas).
 - c. Resultados:
 - i. Ficha de identidad del sitio (y de cada sección principal si es el caso) incluyendo:
 1. Título
 2. Frase descriptiva (tagline)
 3. Metadatos-Description
 4. Metadatos-Keywords
2. Diseño:

- a. Objetivos:
 - i. Optimización de la página principal del sitio y/o de las secciones principales.
 - b. Herramientas:
 - i. Programas de edición de páginas web o sistemas de gestión de contenidos.
 - c. Resultados:
 - i. Código fuente del sitio web optimizado para la campaña en los apartados de:
 - 1. metadatos
 - 2. contenidos
 - 3. texto del sitio
 - 4. navegación del sitio
 - 5. enlaces internos
 - 6. enlaces hacia otros sitios
 - 7. enlaces desde otros sitios
 - 8. etc.
3. Implantación:
- a. Objetivos:
 - i. Desarrollo de la campaña.
 - b. Herramientas:
 - i. Formularios de alta de los distintos directorios y motores; fichas de seguimiento, etc.
 - c. Resultados:
 - i. Acciones de alta.
 - 1. En directorios.
 - 2. En motores de búsqueda
 - ii. Acciones de captación de enlaces hacia el sitio.

Este ciclo de vida propuesto para una campaña *SEO* es del año 2004, por lo que actualmente ha variado ligeramente al aparecer nuevos elementos para promocionar páginas web, como *Google Adwords* o los directorios de artículos, y al haber mejorado los motores de búsqueda sus algoritmos, los cuales ahora dan más importancia al contenido en sí de las páginas web que a los factores externos de las mismas.

Es por eso que **esta investigación está centrada en el estudio del contenido y la estructura de páginas web en relación a los factores de posicionamiento de los buscadores**, porque el *SEO* cada vez va más encaminado a la creación de contenido original y el posicionamiento del mismo por las palabras clave adecuadas, ya no es como hace años que lo más importante era conseguir enlaces externos hacia una página web para posicionar la misma.

2.4 Factores *on-page*

Los factores *on-page* son los factores internos de una página que afectan a su posicionamiento en los buscadores. Estos factores no sólo son referentes a la estructura del código, que son los que van a estudiarse en esta investigación, sino que hay muchos más factores, los cuales aparecen a continuación (Casares 2010; Carlos 2012; Cornello 2010).

2.4.1 Dominio

- Edad del dominio
- Fecha de expiración
- Información de registro (Whois Privacy)
- Dominios de nivel superior y geográficos (foco geográfico, por ejemplo .com y .es)
- Dominio de nivel superior (por ejemplo .com y .info)
- Dominios y subdominios
- Últimos registros del dominio (con qué frecuencia cambia de IP)
- Últimos propietarios del dominio (la frecuencia con que cambia de propietario)
- Palabras clave en el dominio
- IP del dominio
- Vecinos que comparten IP con el dominio (mismo servidor)
- Páginas que mencionan el dominio (sin necesidad de enlazar a él)
- Ámbito de actuación geográfico (Google Webmaster Tools)

2.4.2 Servidor

- Ubicación geográfica del servidor
- Fiabilidad del servidor / tiempo de actividad y caídas

2.4.3 Arquitectura

- Estructura de las URL
- Estructura del código "HTML"
- Estructura semántica (Identificar los elementos de la pantalla, uso de H1-H6...)
- Uso de CSS / Javascript externos
- Accesibilidad de la estructura web (uso de tecnologías accesibles o no, JavaScript, etc)
- Uso de URL canónicas
- Código HTML correcto
- Uso de cookies

2.4.4 Contenido

- Lenguaje utilizado para la redacción de los contenidos
- Contenidos únicos
- Cantidad de contenido respecto al total de las pantallas (% de texto y HTML)
- Densidad de contenidos enlazados (links en lugar de texto)
- Texto en relación al contenido (sin enlaces, imágenes, código, etc)
- Contenidos actuales o puntuales (por ejemplo para búsquedas de temporada)
- Información semántica (indexación basada en la frase así como desórdenes frase indicadores)
- Identificar los tipos de contenidos en la categoría principal (transacciones, información, navegación)
- Contenido / nicho de mercado
- Marcar las palabras clave (juegos de azar, ...)
- Imágenes con texto
- Contenido malicioso (puede ser incorporado por hackers)
- Errores ortográficos, y párrafos sin signos de puntuación
- Utilización de frases originales (no copiadas)

2.4.5 Enlaces internos

- Poner links hacia un contenido específico en una pantalla (#)
- Texto de los enlaces internos (enlazar a una URL siempre con el mismo texto)
- Uso de migas de pan / breadcrumbs / ruta desde el inicio
- Uso de enlaces con "nofollow" (microformatos)
- Densidad de enlaces internos

2.4.6 Elementos del sitio web

- Contenido del archivo robots.txt
- Frecuencia de actualización de la página
- Tamaño de la página (número de pantallas)
- Antigüedad de la página en Google
- Importancia del mapa web XML
- Páginas de confianza (información de contacto (para la búsqueda local, aún más importante), política de privacidad, Condiciones de servicio, Aviso legal y similares)
- Tipo de Web (por ejemplo, blog en lugar de sitios de información en el Top 10)

2.4.7 Factores específicos de cada pantalla

- Tag Robots
- Edad de la pantalla

- Frescura de la pantalla (Frecuencia de ediciones y % de cambios en modificaciones sucesivas)
- Duplicidad de contenidos en diferentes pantallas de la página (contenido duplicado interno)
- Nivel de lectura para una comprensión adecuada
- Tiempo de carga de las pantallas
- Títulos de las pantallas
- Popularidad de la pantalla para enlaces internos
- Popularidad de la pantalla para enlaces externos (enlaces externos que son relevantes para otras pantallas del sitio)

2.4.8 Uso de palabras clave y su prominencia

- Palabras clave en el título de la pantalla
- Palabras clave al principio del título de la pantalla
- Palabras clave en las etiquetas ALT
- Palabras clave en el texto del enlace de enlaces internos (ancla texto interno)
- Palabras clave en el texto de los enlaces salientes
- Palabras clave en textos en negrita y cursiva
- Palabras clave al inicio del texto
- Palabras clave en el cuerpo del texto
- Sinónimos de palabras clave relacionadas con la temática de la pantalla y la página
- Nombrar los archivos con palabras entendibles
- Palabras clave en las URL
- No simular que casualmente la palabra clave está en todas partes (en el dominio, el nombre de archivo, la primera palabra del título la primera palabra de la primera línea de la descripción y etiqueta de palabras clave ...)
- No abusar de palabras clave en etiquetas HTML

2.4.9 Enlaces salientes

- Número de enlaces salientes (por dominio)
- Número de enlaces salientes (por pantalla)
- Calidad del destino de los enlaces salientes
- Enlaces a malos vecindarios
- Relevancia de los enlaces salientes
- Enlaces a páginas 404 y otros errores
- Enlaces desde el sitio web del cliente a empresas SEO
- Imágenes "Hot-linked"

Lo cierto es que la mayoría de estos factores no está demostrado científicamente que afecten al posicionamiento de páginas web, puesto que muchos de ellos no es posible actualmente estudiarlos de manera individual aislándolos del resto para poder comparar resultados.

La lista superior simplemente trata de identificar una serie de características que, si se optimizan, en conjunto mejoran el posicionamiento de páginas web específicas.

2.5 Factores *off-page*

Los factores *off-page* son aquellos que, en teoría, no pueden ser controlados por la página web a posicionarse. Digo en teoría porque, al final, siempre se puede mediando con algunos elementos de la web.

Ninguno de estos factores va a ser analizado en esta investigación, pero sí es necesario saber cuáles son para aislar los factores que queremos estudiar de los que no.

Los principales factores *off-page* de los buscadores son los siguientes (CARLOS 2012):

2.5.1 Enlaces entrantes

- La relevancia de las páginas que enlazan
- La relevancia de las pantallas que enlazan
- Calidad de los sitios que enlazan
- Calidad de las pantallas desde las que se enlaza
- Enlaces de entrada desde redes de sitios
- Enlaces de orígenes similares entre varias páginas
- Variedad en los tipos de enlaces:
- Variedad en los textos del enlace
- Enlaces desde varias IP's
- Diversidad geográfica de los enlaces entrantes
- Enlaces procedentes de diferentes dominios
- Temática de la página de origen del enlace
- Enlaces procedentes de diferentes tipos de páginas (blogs, directorios, etc)
- Situación de los enlaces en las pantallas de origen (lateral, pie ...)
- Autoridad del origen (CNN, BBC, etc) de los enlaces entrantes
- Enlaces desde páginas inapropiadas (la presencia o ausencia de vínculos de retroceso de lugares marcados como malos vecinos)
- Ratio de enlaces recíprocos (en relación con un promedio general)
- Ratio de enlaces procedentes de redes sociales vínculos relación (enlaces de medios sociales en comparación con el promedio general)
- Evolución de los vínculos de retroceso, tendencias y patrones
- Citas en Wikipedia y Dmoz
- Perfil de los registros históricos de enlaces entrantes (compra / venta ..., etc)
- Enlaces procedentes de páginas de marketing social

2.5.2 Enlaces entrantes por separado

- Autoridad del dominio de primer nivel (extensión)

- Autoridad del dominio que enlaza
- Autoridad de la pantalla que tiene el enlace
- Localización del enlace en la pantalla (pie de página, espacio de navegación, cuerpo del texto)
- Texto del enlace (y la etiqueta alt de las imágenes con enlaces)
- El atributo "Title" en un enlace

2.5.3 Perfil de los Visitantes y conducta

- Número de visitas
- Procedencia geográfica
- Porcentaje de abandonos (el rebote)
- Costumbres de navegación de los visitantes (qué otros lugares visitan)
- Tendencias y patrones de los visitantes (puntas repentinas de tráfico en la página)
- Frecuencia con que se pulsan los resultados de la página en las búsquedas (en relación con otros resultados)

Al igual que la lista de factores *on-page* descrita en el apartado anterior, esta lista de factores tampoco está comprobada científicamente por el mismo motivo: hasta la fecha de hoy, no se tiene constancia de que alguien haya aislado factor por factor para estudiar su influencia de forma individual.

En esta investigación no van a medirse ninguno de estos factores *off-page*, ya que el objetivo principal es medir los factores *on-page* mencionados en las hipótesis de forma aislada.

2.6 Buenas prácticas de SEO según Google

Google recomienda una serie de directivas a seguir si alguien quiere que su sitio web esté optimizado para los buscadores. Obviamente, estas directivas no son el padre nuestro del SEO, ya que hay infinidad de cosas que Google no quiere que sepa nadie. Las buenas prácticas que Google recomienda para que una página web sea más accesible para los buscadores son las siguientes (Google 2012):

1. Crear títulos de página únicos y precisos: Según Google, la etiqueta <title> del código HTML es la parte más fundamental de toda la página web, ya que es lo que primero ve el usuario cuando realiza una búsqueda. Es decir, los resultados de búsqueda se muestran por su título.
2. Utilizar la meta etiqueta "description": Es fundamental, ya que como el título, en la mayoría de los casos se muestra como texto debajo del título en las SERPs de los buscadores. Esta es menos importante porque no siempre se muestra la "description", sino que a veces Google toma otras partes del texto de la página web.
3. Mejorar la estructura de las URL de las páginas: Esto es lo que se conoce como "crear URLs amigables", es decir, URLs que contengan palabras clave y que sean entendibles para los buscadores. Para esto Google recomienda crear URLs del tipo "<http://www.misitio.ext/dir1/dir2/texto-descriptivo-con-palabra-clave/>", de forma

que no haya más de dos subdirectorios en la URL y el texto, en caso de necesitar ser separado, se utilicen guiones. Además, recomienda evitar URLs del tipo “<http://www.misitio.ext/index.php?id=4&order=name>”, ya que este tipo de URL no aporta ninguna información al buscador.

4. Crear una buena estructura de navegación en el sitio: Esto es importante para ayudar a los buscadores a clasificar la información de un sitio web. Hacer un estudio de las categorías de contenido que van a emplearse, así como de las tags, es fundamental. En caso de que sea una página de venta de productos, también es importante la organización de los casos de uso para saber cómo tenemos que decirle a los buscadores que queremos que entren los usuarios a nuestra web.
5. Crear Sitemap: Los mapas del sitio ayudan a los buscadores a indexar todas las páginas de un sitio web, ya que son páginas a través de las cuales hay enlaces a todo el sitio. Además, tener un fichero sitemap.xml en el servidor y enseñárselo a Google también ayudará a que este indexe páginas más rápidamente.
6. Ofrecer contenido de calidad: Esto está tomando especial importancia en los últimos tiempos. A los buscadores les gusta el contenido fresco y actualizado. Además, les gusta el contenido único, por lo que Google penaliza el contenido duplicado o copiado de otras páginas web, y tiene formas de saber qué web ha sido la que ha publicado primero un contenido.
7. Textos ancla de los enlaces: Para Google, un enlace a una página web, ya sea dentro o fuera de un sitio web, es un voto de confianza para esa página siempre que el enlace sea de calidad, es decir, el sitio que enlaza esté relacionado con el enlazado y el texto ancla del enlace sea adecuado. Por eso cobra especial importancia pensar qué palabras se insertarán en los textos ancla o “anchor text” de los enlaces de nuestro sitio.
8. Imágenes: Las imágenes también son contenido para los buscadores. El nombre del fichero de la imagen es empleado por los buscadores, así como el atributo “alt” de la etiqueta , para tratar de entender qué es lo que muestra la imagen. Por ello es importante describir bien las imágenes, tanto sus nombres de fichero como sus textos alternativos.
9. Encabezados: Dar una buena estructura al documento empleando los encabezados <h1>, <h2>, ... , <h6> es muy importante, así como incluir palabras clave dentro de los mismos (sin abusar).

Hay más prácticas que Google recomienda, pero tienen más que ver con la no accesibilidad de los buscadores a sitios, así como otras pautas orientadas a sitios web para móviles que no nos interesan en esta investigación.

Google menciona que estas prácticas son recomendadas, no da ningún dato científico de que de verdad mejoren el posicionamiento de páginas web. La idea de realizar esta investigación surge principalmente de este documento difundido por el que hoy es uno de los buscadores más utilizados en el mundo, y el objetivo es conseguir medir cómo estas buenas prácticas afectan al SEO aportando números capaces de construir una métrica.

Además, algunas de estas prácticas que menciona Google son bastante genéricas, y consideramos que pueden ser más específicas. Por ejemplo, en el documento se menciona que es importante cómo se construyen las etiquetas de título en referencia a

si se incluyen o no palabras clave en ellas. Pero se puede ir más allá, ya que puede ser que además, la posición de la palabra clave en el título también haya que tenerla en cuenta.

2.7 Técnicas SEO de sombrero blanco y técnicas SEO de sombrero negro

También conocidas como “*White Hat SEO*” y “*Black Hat SEO*” respectivamente.

En el mundo del SEO existen muchas técnicas y muy variadas técnicas que personas que se dedican a posicionar páginas web en los motores de búsqueda utilizan. Desde hace mucho tiempo estas técnicas, además de otras clasificaciones, se dividen en dos grupos en relación a si son buenas o malas prácticas.

Las buenas prácticas SEO son denominadas técnicas “*White-hat SEO*” o técnicas de sombrero blanco, mientras que las técnicas que son fraudulentas para los buscadores son denominadas como “*Black-hat SEO*”.

Entre las primeras técnicas encontramos, básicamente, todas las buenas prácticas que Google nos recomienda en su libro para principiantes en SEO de buenas prácticas para diseñar páginas web amigables para los motores de búsqueda(Google 2012).

Algunas de estas técnicas ya han sido mencionadas más arriba, como hacer hincapié en el título de la página, generar meta descripciones diferentes en cada página web, hacer un uso correcto de los encabezados...(Google 2012)

En cuanto a las malas prácticas o “*Black-hat SEO*”, encontramos las siguientes (solo se mencionan técnicas *on-page*):

1. Cloaking: Esta técnica consiste en mostrar una página web diferente a los buscadores que a los usuarios, de forma que a los motores de búsqueda se les muestra una página sobreoptimizada para ellos cuando acceden a una determinada página web, mientras que a los usuarios se les muestra otra diferente más orientada a la usabilidad. Con esto se consigue optimizar la web para buscadores y personas (algo difícil de conseguir)(López Gómez 2009)
2. Texto invisible: Esta técnica mal vista por los buscadores consiste en colocar texto en la página web del mismo color que el fondo del lugar en el que se encuentra. De esta forma, el texto es invisible para los usuarios mientras no lo marquen pero es totalmente visible para los buscadores. Normalmente en este tipo de texto lo que se incluía eran palabras clave para posicionar de forma desmesurada.
3. Duplicación de dominios: Otra técnica muy utilizada tiempo atrás, que consistía en redirigir infinidad de dominios a una misma página web para copar todos los resultados de búsqueda de una determinada clave.
4. Enlaces forzados: También llamados “spam keywords”. Técnicas que lo que intenta es, a través de la creación de decenas o a veces incluso cientos de enlaces en una misma página web, posicionar la misma por las palabras clave de estos enlaces u otras páginas web del mismo sitio.

Estas son sólo algunas de las técnicas no recomendadas más conocidas en el mundo del SEO. No son recomendadas porque los algoritmos de los motores de búsqueda, especialmente Google, ya son capaces de reconocerlas desde hace tiempo, por lo que penalizan las páginas web que son detectadas utilizando estas malas prácticas(López Gómez 2009).

En esta investigación se tienen en cuenta estas técnicas y se generan páginas web que no hacen uso de ellas para que esto no interfiera en la medición de los resultados de la investigación.

2.8 Ingeniería inversa y SEO

Actualmente, además de fiarnos de lo que nos dicen los ingenieros de los principales motores de búsqueda, la única forma de averiguar cómo funcionan estos es a través de la ingeniería inversa, tal y como se explica en uno de los artículos referenciados al final de este documento(Zhu and Wu 2011).

En este artículo, lo que hacen sus autores para medir diferentes factores de la optimización de páginas web para buscadores es básicamente lo que se realizará en este proyecto de investigación, pero con algunos matices importantes:

1. Estudian factores *on-page* y *off-page* a la vez, con lo que realmente no se obtienen conclusiones claras de cuáles son los que están afectando realmente al posicionamiento de cada página.
2. Basan su estudio en páginas web que ya están colgadas en Internet y que son casos reales, y además lo dividen en categorías diferentes en función del PageRank(Carlos 2012).

Estos dos matices son importantes ya que **el estudio de esta investigación se basa en un conjunto de páginas generadas aleatoriamente con contenidos que de forma segura están orientados a posicionar una palabra clave**, mientras que de otra forma no podemos asegurar esto. Además, el objetivo principal de esta investigación está centrado únicamente en ciertos factores *on-page* del posicionamiento, mientras que en el mencionado artículo se estudian todos simultáneamente.

2.9 Ideas adoptadas del estado del arte

Durante la lectura de los diferentes artículos, libros y referencias del estado del arte mencionado en las referencias se han podido adoptar una serie de ideas que han ayudado a intentar mejorar la forma en que van a obtenerse los resultados de esta investigación, tanto en la preparación del entorno como en la metodología en sí.

2.9.1 Artículo “Cómo mejorar tu ranking en Google: Mitos y realidades”

En este artículo(Su et al. 2010) los autores hacen básicamente lo que se realizará en esta investigación: toman una serie de factores que comúnmente se cree que pueden afectar al posicionamiento de las páginas web en los buscadores, como la etiqueta de título, la autoridad de un dominio, los encabezados, densidad de la palabra clave a posicionar... todos ellos factores *on-page*, al igual que en este trabajo.

Sin embargo, a pesar de estar estudiando únicamente factores *on-page* de las páginas web, se basan en muestras que ya existen en Internet desde hace tiempo, y esto tiene el inconveniente de que, casi con un 100% de probabilidad, sobre esas páginas web están influyendo factores externos, como el número de *backlinks* que apuntan hacia ellas o el texto de estos.

Esto claramente distorsiona los resultados de la investigación ya que, según ellos mismos afirman, han sido capaces de simular el motor de ordenación de Google para el 80% de los casos, es decir, su algoritmo en teoría se parece en un 80% al de Google(Su et al. 2010).

No obstante el otro 20% queda en el aire, y es ahí donde es mejorable la investigación. En este trabajo lo que se prioriza es aislar los factores internos de la página de los externos, de tal forma que solo se estudiarán los primeros, pero asegurándonos de que los resultados son 100% reales. Es decir, ante dos páginas que tienen cero factores externos influyéndoles, nosotros seremos capaces de discernir cuál de las dos quedaría mejor posicionada.

Así, quedaría como trabajo futuro estudiar los factores externos de forma aislada, aunque esto es más complicado debido al problema del contenido duplicado, pero ese es otro tema a tratar más adelante.

2.9.2 Matices de la investigación

La mayoría de todas las investigaciones, libros, artículos y referencias en Internet consultadas tienen algo en común: no tratan de aislar los factores a estudiar. Algunas solo estudian un tipo de factores pero sin aislarlos, y otras directamente intentan estudiar todos los factores a la vez.

Esto, debido a la naturaleza de Internet es prácticamente imposible, ya que es una red tan grande que prácticamente cualquier contenido existente está duplicado varias veces y enlazado por cientos de páginas web, factores que podrían influir en las posiciones que ocupan las páginas web en las búsquedas.

Todos los artículos consultados hablan de sofisticados algoritmos desarrollados para imitar al de Google en un porcentaje de búsquedas o de obtención de muestras de los buscadores para posteriormente analizarlas directamente.

En el primer caso, intentar imitar un algoritmo que lleva más de 20 años desarrollándose es poco menos que una quimera y, además, se deja de lado el resto de motores de búsqueda.

En el segundo caso, el que en mi opinión es el mejor escogido, se trabaja con muestras de consultas a los buscadores que tienen factores externos que les afectan, sin haber desarrollado ellos mismos sus propias muestras.

Para el desarrollo de esta investigación hay varios puntos que se tendrán en cuenta:

- Aislar factores internos de las páginas o factores *on-page*.
- Intención de estudiar varios buscadores: Google, Yahoo y Bing.
- Generar muestras propias a partir de textos de muestra, las cuales no tendrán sentido semánticamente hablando. Cada párrafo tendrá sentido por sí mismo y será ortográficamente y gramaticalmente correcto, pero el documento en su totalidad no tendrá sentido semánticamente hablando a pesar de estar orientado a una palabra clave.

PARTE III - DESARROLLO DE LA INVESTIGACIÓN

Capítulo 3. Desarrollo de la investigación

El desarrollo de la investigación es bastante lineal, ya que para realizar un paso siempre se requieren todos los anteriores, por lo que explicar la línea seguida es bastante sencillo una vez se ha estudiado el estado del arte mencionado en el capítulo anterior:

1. Identificar los factores que van a estudiarse en la investigación en base al estado del arte estudiado: realmente este punto ya se ha realizado pero en este capítulo se explicará el porqué de cada uno con más detalle.
2. Técnica para aislar los factores a estudiar.
3. Identificar los motores de búsqueda sobre los que se va a experimentar.
4. Identificar los riesgos que conlleva la investigación: principalmente es uno, el cual está explicado más adelante en este capítulo.
5. Crear el generador de páginas web.
6. Crear diferentes muestras de páginas web empaquetadas por palabras clave listas para ser estudiadas en los diferentes buscadores.
7. Conseguir indexar las diferentes muestras en los buscadores seleccionados: aquí es donde reside el principal riesgo, que los buscadores no indexen las páginas web por algún motivo.
8. Pautas a seguir para analizar los resultados: cómo va a calcularse la métrica que siguen los buscadores en base a los resultados obtenidos en las SERPs.
9. Una vez conseguimos que las muestras hayan sido indexadas, se realizan búsquedas por las palabras clave de cada muestra para estudiar los resultados devueltos por cada buscador.

3.1 Factores a estudiar en la investigación

Como ya se ha comentado anteriormente en diferentes apartados, hay diferentes factores que afectan al posicionamiento de una página web en los buscadores para una determinada palabra clave, los cuales pueden dividirse principalmente en dos grupos: factores *on-page* y factores *off-page*. En esta **investigación ha decidido estudiarse el conjunto de factores on-page** (los que son internos y dependientes de la propia página web) porque son los únicos que, en teoría, pueden ser alterados a gusto del creador de la página web (Zhu and Wu 2011).

Los factores *off-page*, determinados principalmente como es sabido y promovido por Google desde hace muchos años, vienen marcados principalmente por el número de enlaces que apuntan a una determinada página web (Casares 2008).

En teoría este número de enlaces que apuntan a una página web no podría ser alterado por el creador de la misma, aunque en la práctica obviamente esto no es así, ya que existen técnicas para ello, las cuales se salen de esta investigación y podrían ser un trabajo futuro que no compete el objetivo de la investigación.

En realidad, se sabe que existen factores *on-page* que afectan al posicionamiento de una página y se intuye cuáles pueden ser, pero no existen evidencias científicas ni afirmaciones directas por parte de los buscadores de cuáles son estos factores.

En base a todas estas intuiciones y creencias van a tomarse para estudio una serie de factores en esta investigación y, a partir de estos factores, se intentará obtener una métrica para intentar averiguar el peso de cada uno de ellos. Puede ser que el peso sea nulo, con lo cual ese factor no será relevante, o puede que no. De todos esos factores, es muy difícil que alguno no afecte realmente al posicionamiento de las páginas web, ya que por sentido común y analizando las *SERP (Search Engine Result Pages)* hay indicios para algunos de ellos, por lo que es seguro que la investigación devolverá algún resultado concluyente.

De todos los factores *on-page* que se cree que afectan a la posición de una página web para una determinada palabra clave, para esta investigación se han dividido en dos tipos:

- Los que son intrínsecos del código HTML: Obviamente no es descabellado pensar que, si el contenido de una página web afecta al posicionamiento de la misma (ya que el posicionamiento se realiza en base a las palabras clave que aparecen en la página), también es importante en que partes del código HTML aparece mencionada una palabra clave. Por ejemplo, se cree que si la palabra clave aparece entre las etiquetas de título, esta es muy importante, mucho más que si no apareciera.
- Los que no dependen del código: Hay muchas opiniones en la red y en diferentes artículos y libros acerca de estos factores, aunque no se consideran tan importantes como los primeros. Podríamos agrupar dentro de estos posibles factores, por ejemplo, la ubicación de la IP donde se aloja el servidor de la página web (para las búsquedas locales), el tiempo de carga de una página web o, sin ir más lejos, el idioma de la misma. Se sabe que este tipo de factores existen porque si por ejemplo entras en <http://google.es> (Google España) y realizas una búsqueda no devuelve los mismos resultados que si se realiza desde <http://google.co.uk> (Google United

Kingdom), y lo vamos a demostrar con una imagen: buscamos por la palabra clave “Smartphone” en ambas URL de Google y vemos los resultados.

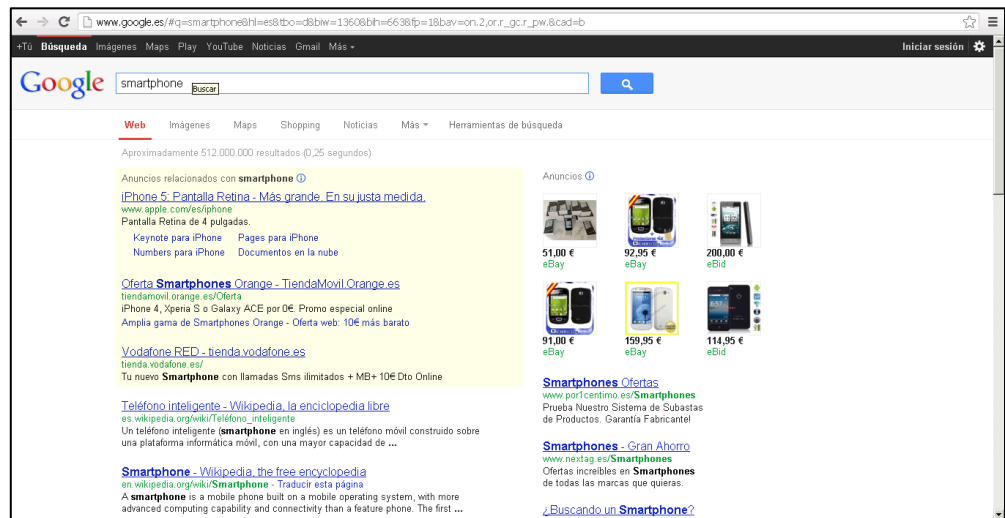


Ilustración 6 – Resultados para la palabra clave “Smartphone” en <http://google.es>

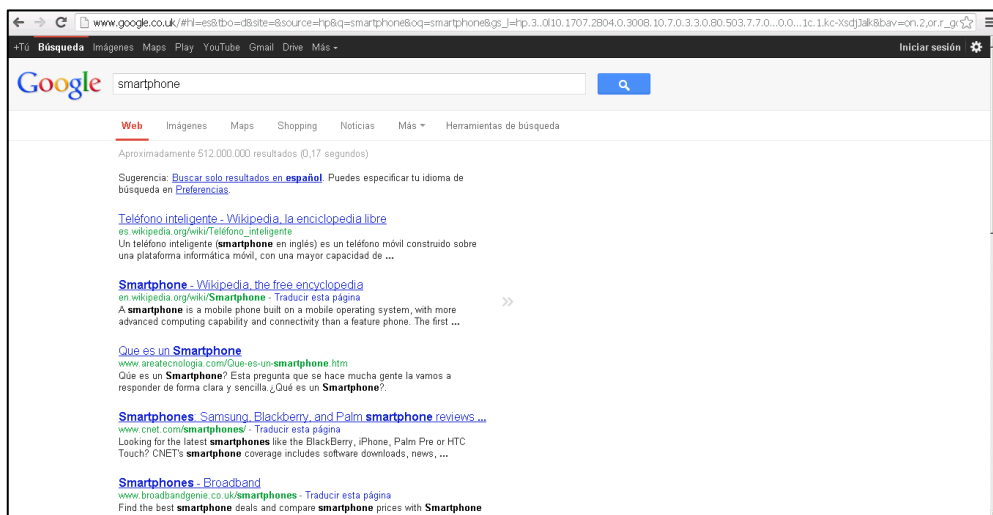


Ilustración 7 – Resultados para la palabra clave “Smartphone” en <http://google.co.uk>

Todos estos supuestos factores han sido comentados en el apartado del estado del arte y aparecen recopilados en una de las referencias(Carlos 2012).

Para esta investigación en concreto únicamente van a estudiarse el primer subtipo de factores dentro de los factores *on-page*: los factores intrínsecos al código HTML de la página. Otro posible trabajo futuro sería estudiar el resto de factores dentro de este tipo, los que son dependientes de la página web pero no del código HTML.

Los **factores seleccionados para su estudio** vienen resumidos en las hipótesis de la investigación, dentro del apartado de la introducción, pero aquí los desglosaremos con más detalle:

- Que la palabra clave aparezca en el título de la página: Este parece ser uno de los factores más importantes a juzgar por la apariencia que tienen las SERP. Cuando se

realiza una búsqueda por una palabra clave, la misma casi siempre aparece en el título mostrado por el buscador en los primeros resultados devueltos tal y como podemos ver en las siguientes imágenes:

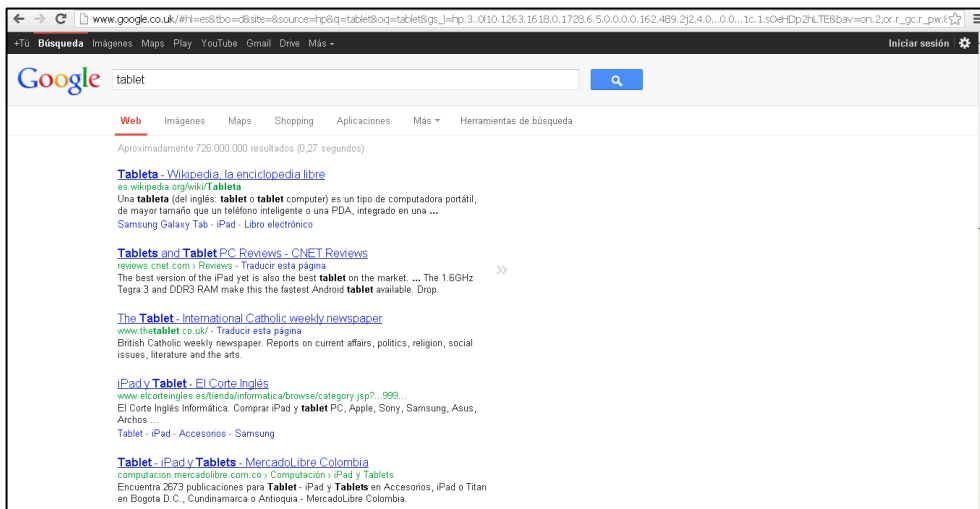


Ilustración 8 – Resultados de búsqueda de la palabra clave “Tablet” en Google

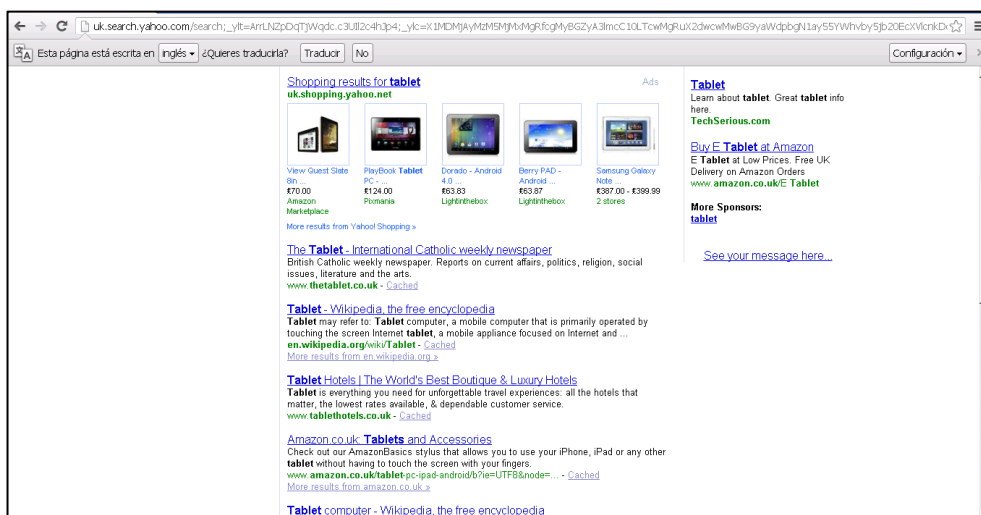


Ilustración 9 – Resultados de búsqueda de la palabra clave “Tablet” en Yahoo omitiendo los resultados publicitarios

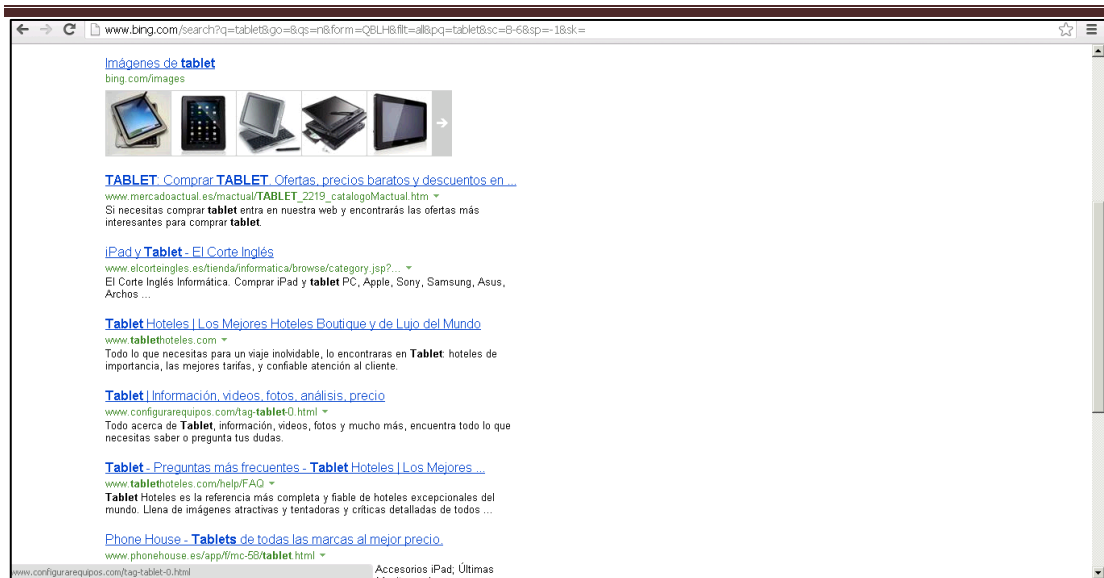


Ilustración 10 – Resultados de búsqueda de la palabra clave “Tablet” en Bing omitiendo los resultados publicitarios

- El número de apariciones de la palabra clave en el título: Este factor está menos claro, pero viendo los mismos resultados que para el factor anterior, vemos que en algunas páginas de las mejor posicionadas la palabra clave aparece dos veces en el título, por lo tanto también se considerará esto para el estudio de esta investigación.
- Que la palabra clave aparezca en la meta-descripción de la página: Desde siempre se ha tenido la creencia de que los *meta-tags* que es posible incluir en el código HTML tienen especial importancia para los buscadores. El texto que se asigna a la meta-descripción es, normalmente, el que los buscadores emplean para describir la página web en las SERP salvo en alguna excepción que pueden emplear texto del propio contenido de la web. Es por ello que este factor puede tener especial importancia para los buscadores a la hora de posicionar las páginas web. Además, tal y como vemos en las imágenes utilizadas para el primer factor, la palabra clave aparece en la mayoría de *meta-descripciones* de las páginas web de los primeros resultados.
- Que la palabra clave aparezca en la meta-title de la página: Por el mismo motivo que en el apartado anterior este puede ser un factor a tener en cuenta para los buscadores, aunque también existe la creencia de que el *meta-title* no sirve para nada ya que los buscadores dan prioridad a la etiqueta de título de la página antes que a esta *meta-etiqueta*.
- Que la palabra clave aparezca en las meta-keywords de la página web: Como en los dos anteriores casos, este puede ser un factor a tener en cuenta. Aunque últimamente existe la creencia de que las palabras clave incluidas en las meta-tags ya no son importantes para los buscadores, en especial para Google (Cutts 2011a), conviene corroborar esta creencia.
- Que la palabra clave aparezca en la etiqueta H1: Este se cree que también es un factor muy importante ya que, dentro de la estructura del código HTML, esta

etiqueta se emplea normalmente para describir al usuario que puede encontrar en la misma en líneas generales.

- Que la palabra clave aparezca en la etiqueta H2: Al igual que la etiqueta H1, esta tiene también importancia para el usuario, ya que, al igual que sus sucesoras (H3, H4, H5, H6), ayudan a clasificar la información de la página web y esquematizarla, algo que es útil para que los buscadores sepan qué contenido hay en una página web en teoría.
- Que la palabra clave aparezca en la etiqueta H3: Lo mismo que la anterior. Las sucesivas etiquetas de H4 en adelante no van a estudiarse porque sería redundante.
- Que la palabra clave aparezca en la primera frase de la página web: Se cree que los buscadores dan mucha importancia al principio del contenido de una página web y que, a medida que se desciende en la página, las palabras tienen menor relevancia, por lo que, en teoría, situar la palabra clave en la primera frase o en el primer párrafo ayudaría a posicionar una página web por esa palabra clave.
- Que la palabra clave aparezca en la última frase de la página web: Si en el anterior elemento de la lista decíamos que a medida que se descendía en el contenido de la página web se perdía importancia para las palabras que aparecían, para esto se cree que hay una excepción: la de la última frase o párrafo de la página. ¿Por qué? Porque normalmente en esta parte del contenido suele aparecer un resumen del mismo y esto sería de mucha ayuda para los buscadores a la hora de indexar una página web por una palabra clave.
- El número de veces que la palabra clave aparece resaltada en negrita, cursiva o subrayada: Puesto que estas tres modificaciones del formato hacen resaltar al usuario una determinada parte del texto, se cree que los buscadores analizan esto para intentar obtener las partes importantes de un texto dentro de una página web. Recordemos que, como se dijo en la introducción del estado del arte, **el SEO es ayudar a los buscadores a indexar contenido**, y que mejor forma de hacerlo que marcando lo que es importante. Obviamente si el texto resaltado con estas modificaciones es demasiado extenso esto no ayudaría.
- El número de veces que la palabra clave aparece en enlaces externos a la web: Los buscadores, desde el inicio de Google, ven Internet como un conjunto de páginas web entrelazadas entre sí. En esencia eso es lo que es Internet en realidad. Y por ello, en teoría, una página web que enlaza a otra, es porque ambas tienen algo en común. Por tanto para los buscadores sería importante estudiar los enlaces externos que hay en una página web, para ver el contenido de las páginas enlazadas y que este coincida con la página de partida y el texto del enlace.
- La densidad de la palabra clave en la página: Buscando un poco de información acerca de este posible factor se puede ver fácilmente que hay mucha controversia acerca del mismo. Se cree que una densidad de palabra clave adecuada ayuda mucho al posicionamiento de una página web pero, ¿Cuál es esa densidad?(Cutts 2011b) Se barajan muchas cifras y estudiar una en concreto sería absurdo, por lo que se estudiará el factor en general.
- Que la palabra clave aparezca en el texto alternativo de las imágenes: Este podría ser otro factor a tener en cuenta por los buscadores, ya que normalmente las

imágenes que aparecen cerca del texto del contenido de una página web describen el mismo.

- Que la palabra clave aparezca en la URL de la página: Puesto que en las SERP de los motores de búsqueda la URL es uno de los datos que se muestra al usuario, cabe pensar que los buscadores emplean también la URL para clasificar las páginas.

El punto en el que precisamente fallan las investigaciones acerca de estos factores es que no aíslan estos factores de otros que puedan influir en la obtención de la relevancia de una página web, mientras que **durante esta investigación se aislarán los factores a estudiar del resto de factores para que no haya ninguna alteración en las muestras de resultados que devuelven los buscadores.**

3.2 Técnica para aislar los factores a estudiar

La mejora fundamental de este trabajo de investigación, además de aportar datos numéricos sobre los factores a estudiar, es que dichos factores estarán aislados del resto que no se tienen en cuenta en el estudio.

Cabe recalcar esto porque es algo que no ocurre en los artículos y libros estudiados en el estado del arte.

Para aislar los factores que nos interesan del resto hay que tener en cuenta varias cosas:

1. Sólo nos interesan el tipo de factores *on-page*, es decir, los que son exclusivos de la propia página web y no del exterior. Es por eso que, para evitar que factores externos intervengan, no se creará ningún enlace desde otro dominio o IP hacia las páginas a estudiar. De esta forma nos aseguramos que los enlaces externos no interfieren en los resultados de búsqueda, que son el principal factor externo que puede afectar a los resultados de búsqueda.
2. Dentro de los factores *on-page*, solo nos interesan los factores del código HTML. Se cree que factores como la IP donde se aloja la web, el tiempo de carga, el país del servidor donde se aloja o la autoridad del dominio son factores clave a la hora de posicionar una página web (Carlos 2012), pero como no queremos tenerlos en cuenta basta con alojar todas las páginas a estudiar en el mismo servidor para que estos no interfieran en los resultados.
3. Un factor que últimamente se cree que los buscadores tienen en cuenta es el tema del contenido duplicado (Martínez 2012). Es decir, si dos páginas tienen el mismo contenido, en teoría la que ha copiado el contenido de la otra es penalizada. Para que esto no afecte a los resultados de la investigación es necesario crear muestras de páginas web para una palabra clave con textos muy variados y diferentes estructuras, donde el algoritmo para ello y los archivos de donde se extraerán los textos para cada etiqueta jugarán un papel fundamental.
4. Otra cosa importante a tener en cuenta para los buscadores se cree que es el número de enlaces internos entre páginas web y las palabras que llevan estos en su texto de enlace o *“anchor text”*, por lo que hay que forzar a que todas las páginas web a estudiar tengan el mismo número de enlaces entrantes con las mismas palabras en el texto del enlace. Para ello solo se va a realizar un enlace a cada

página a estudiar desde la misma página web, que será un índice a través del cual los buscadores podrán indexar las páginas web a estudiar. Este enlace tendrá las mismas palabras para todas las páginas de una misma muestra. Un ejemplo para una página de índice podría ser el siguiente:

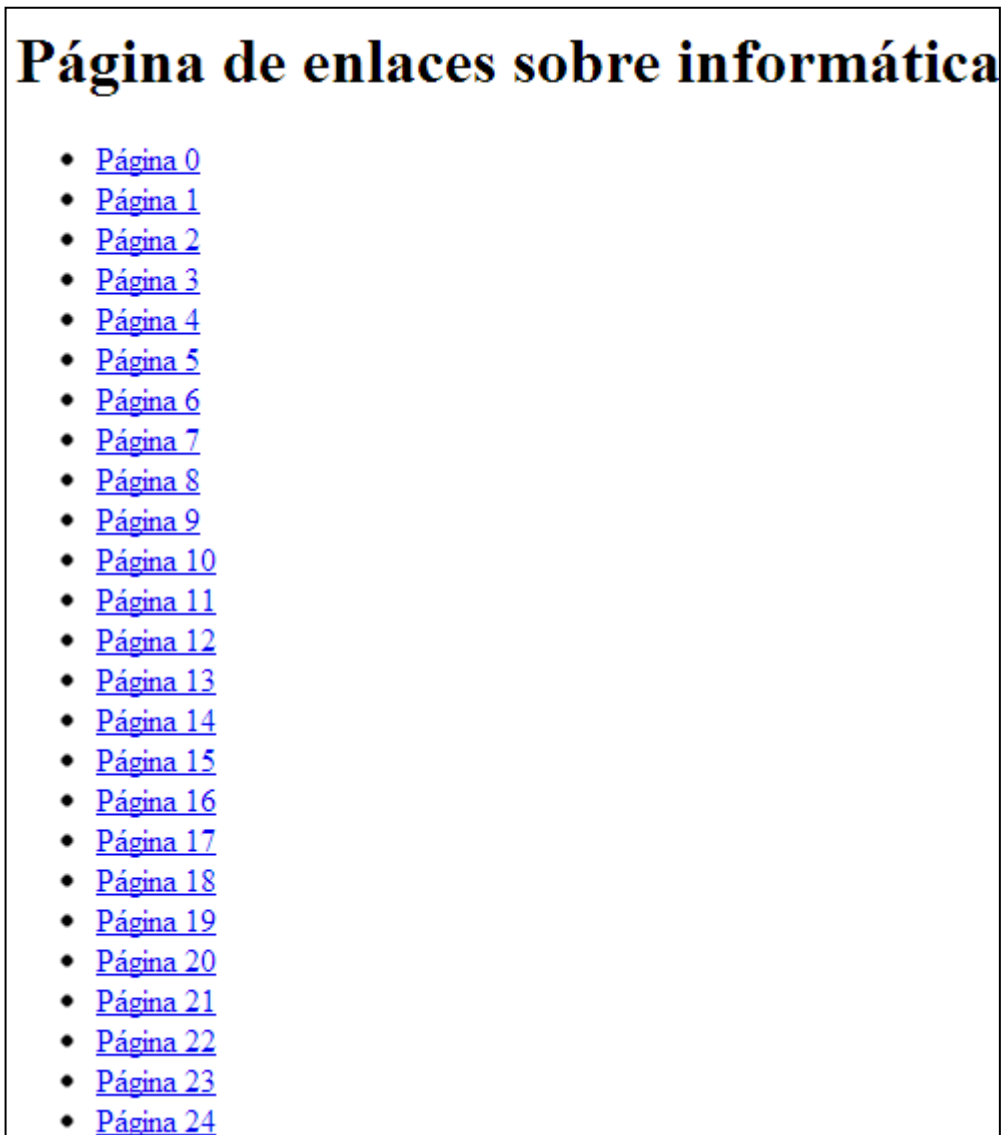


Ilustración 11 – Ejemplo de página de índice para una muestra de páginas web de la palabra clave “informática”

3.3 Motores de búsqueda seleccionados

Según las últimas cifras conocidas, los buscadores más utilizados en Internet son, por este orden, Google, Yahoo y Bing. Google alcanza la mayor cuota de mercado puesto que, en total, el 83% de las búsquedas en Internet mediante PC se realizan a través de este buscador. Yahoo ocupa el segundo lugar con una cuota de un 8%, mientras que Bing tiene una cuota en 2012 de casi el 5% (MarketShare 2012).

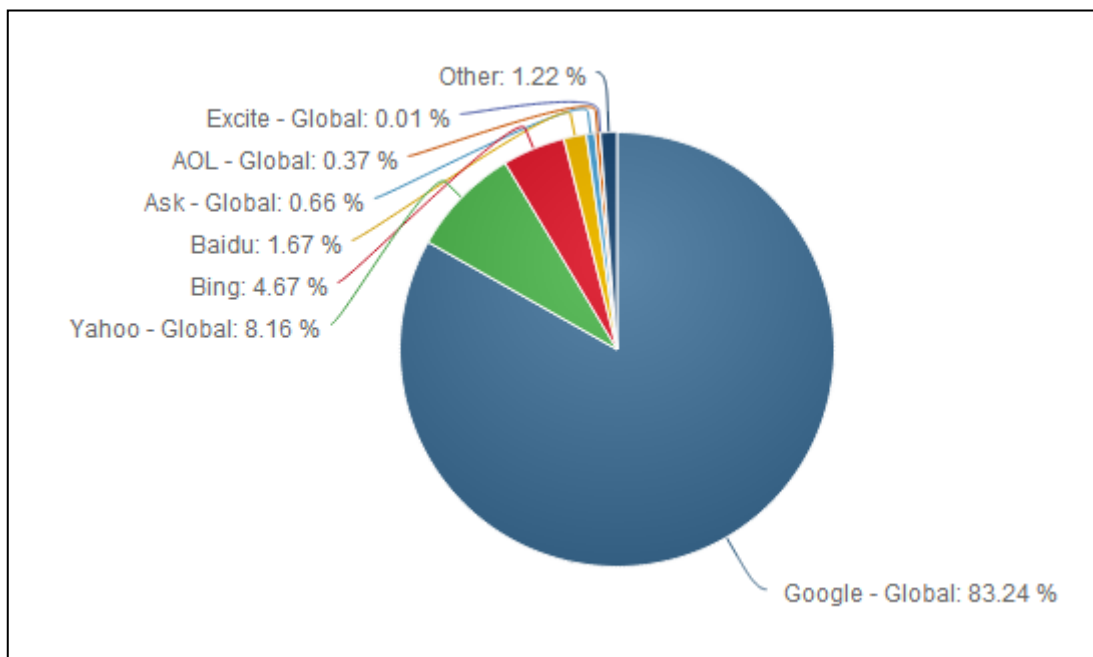


Ilustración 12 – Estadísticas de uso de motores de búsqueda en PC durante el año 2012

Teniendo en cuenta estos datos, los buscadores sobre los que se va a experimentar a lo largo de este trabajo de investigación van a ser estos tres: Google, Yahoo y Bing.

3.4 Riesgos del proyecto de investigación

Esta investigación tiene asociados, además de los riesgos genéricos para todo proyecto de desarrollo o investigación, los siguientes riesgos específicos:

- Que los buscadores no indexen las muestras.
- Que los buscadores indexen tarde las páginas web de las muestras.
- Que los resultados no revelen ningún patrón.

3.4.1 Que los buscadores no indexen las muestras

El principal riesgo que tiene esta investigación es que los buscadores no indexen las páginas web de las muestras a estudiar. Ante este riesgo, si finalmente sucede, un plan de contingencia sería realizar el estudio sobre muestras de páginas web reales. Sin embargo, si obtenemos así los resultados, no nos aseguramos de que solo se tienen en cuenta los

factores que queremos en la investigación; de hecho lo más probable es que se incluya algún otro factor, principalmente el de los enlaces externos hacia las páginas web estudiadas.

En caso de tener que emplear este método, deberíamos buscar una muestra de páginas web en Internet la cual tenga el menor número posible de enlaces externos reconocidos por los buscadores.

Google dispone de una herramienta para saber el número de enlaces externos aproximados que apuntan hacia una página web. Para ello, tan solo debemos insertar en el cuadro de búsqueda un texto con la siguiente estructura: “link:URL”. Así es como obtendremos las páginas web que apuntan a una URL determinada.

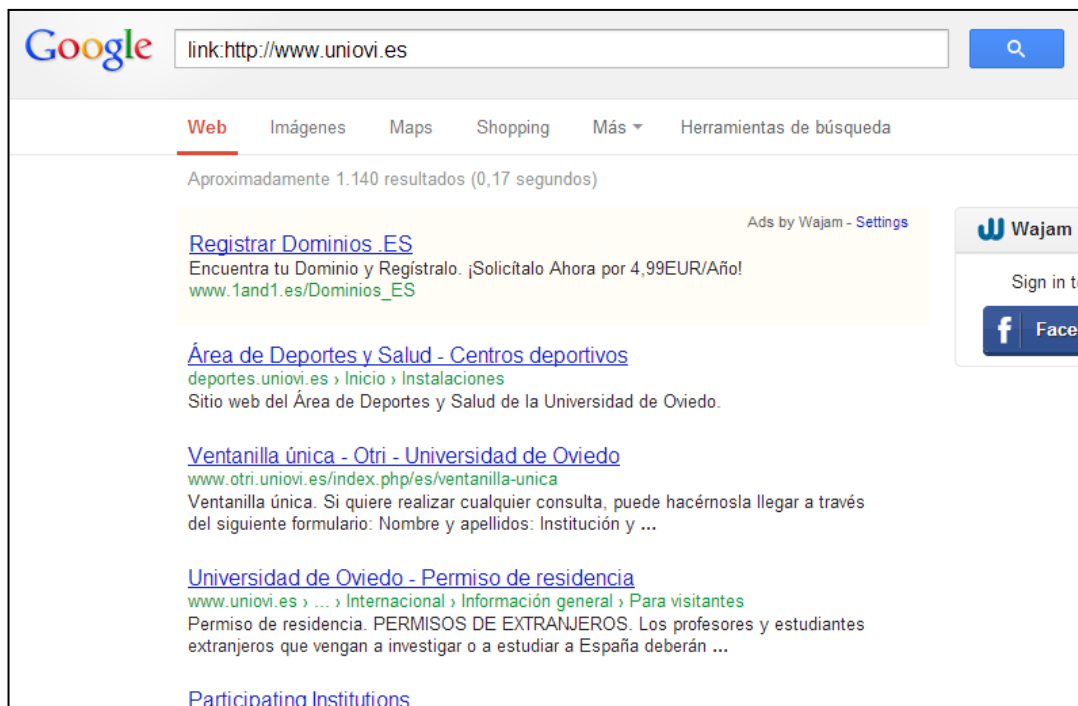


Ilustración 13 – Cómo saber los enlaces que apuntan a http://www.uniovi.es

Para que el factor de los enlaces externos tenga el menor efecto sobre los resultados obtenidos debemos buscar una muestra de páginas web en la que todas ellas apenas tengan menos de 10 enlaces externos, o que todas ellas tengan un número de enlaces apuntando hacia ellas aproximado.

3.4.1.1 No indexación de páginas web por los buscadores

El motivo principal es que un buscador detecte que las páginas son granjas de enlaces, tienen contenido muy similar o duplicado o que, simplemente, la araña de los buscadores encargada de indexar las páginas web tarde demasiado en indexar todas o no sea capaz de llegar a todas las páginas web que necesitamos.

Este último caso es altamente improbable, pero en los dos primeros sí que se pueden tener problemas en esta investigación, por los siguientes motivos:

1. Para que los buscadores puedan acceder a todas nuestras páginas es necesario que exista un enlace a las mismas. Puesto que sería muy complicado crear enlaces desde otras páginas a todas ellas y afectaría además a la autenticidad de los resultados, lo que se hará será crear páginas “índice”. Estas páginas enlazarán a todas las páginas de una muestra y solo será necesario crear un enlace hacia estas, facilitando mucho la labor y no interviniendo así el factor de los enlaces que apuntan hacia páginas web en los resultados de la investigación.
2. Debemos generar muestras basadas en una misma palabra clave para cada caso de estudio, por lo que los textos van a ser, cuanto menos, parecidos. Es por esto que corremos el riesgo de que los buscadores interpreten que alguna página de las que creamos sea copia de otra y no la tenga en cuenta en sus resultados. Esto es muy común especialmente en Google, que tiende a ocultar en las SERP las páginas con contenido similar al resto que ha indexado más tarde en el tiempo. Para evitar esto debemos asegurarnos que tenemos textos de muestra muy variados con los que generar páginas web optimizadas para una misma palabra clave.

3.4.2 Que los buscadores indexen tarde las páginas web de las muestras

Otro riesgo que corre la investigación, aunque menos probable, es que los buscadores indexen las páginas web pero lo hagan tarde. Esto no suele ser habitual ya que actualmente las arañas de los buscadores recorren la red cada 24 o 48 horas, sin embargo esto no era así hace unos años, cuando los buscadores podían tardar hasta más de 1 mes en indexar una página web.

Si finalmente ocurriese esto tan solo quedaría o retrasar el plazo de conclusión de la investigación y su defensa o emplear el mismo método descrito para el riesgo anterior.

3.4.3 Que los resultados no revelen ningún patrón

Uno de los mayores riesgos de una investigación es que una vez concluida esta se llegue a la conclusión que la hipótesis de partida era incorrecta.

En este trabajo de investigación, el riesgo es que una vez realizado el estudio de todos los factores seleccionados, no sea posible identificar un patrón que indique por qué unas páginas web se posicionan mejor que otras.

3.5 Prototipo del generador de páginas web

Para realizar esta investigación se necesitan muestras de páginas web agrupadas por palabra clave. A partir de estas muestras es sobre las que se realizarán búsquedas en los diferentes buscadores y sobre las que se extraerán resultados acerca de patrones que identifican los motores de búsqueda para ordenar los resultados de las SERP.

Es muy importante el método que se empleará para generar las muestras y la calidad de las mismas, puesto que de la calidad de estas muestras y de las páginas web que contienen dependerán los resultados devueltos por los buscadores y el éxito o fracaso final de la investigación.

Para generar las muestras se ha optado por desarrollar un pequeño prototipo en forma de aplicación Java de escritorio que, a partir de unos determinados valores de entrada y unos textos de muestra, es capaz de generar páginas web HTML y, además, un fichero extra para cada una con algunos de sus metadatos, lo cual será de extrema utilidad al final de la investigación para identificar patrones en estos metadatos.

Los buscadores no tendrán constancia de estos ficheros, únicamente se utilizarán para analizar los resultados que ofrecen en las SERP, de forma que buscando el fichero que contiene los metadatos de cada página obtendremos en un rápido vistazo los factores que cumple o no esa página web.

3.5.1 Arquitectura del prototipo

3.5.1.1 Diagrama de paquetes

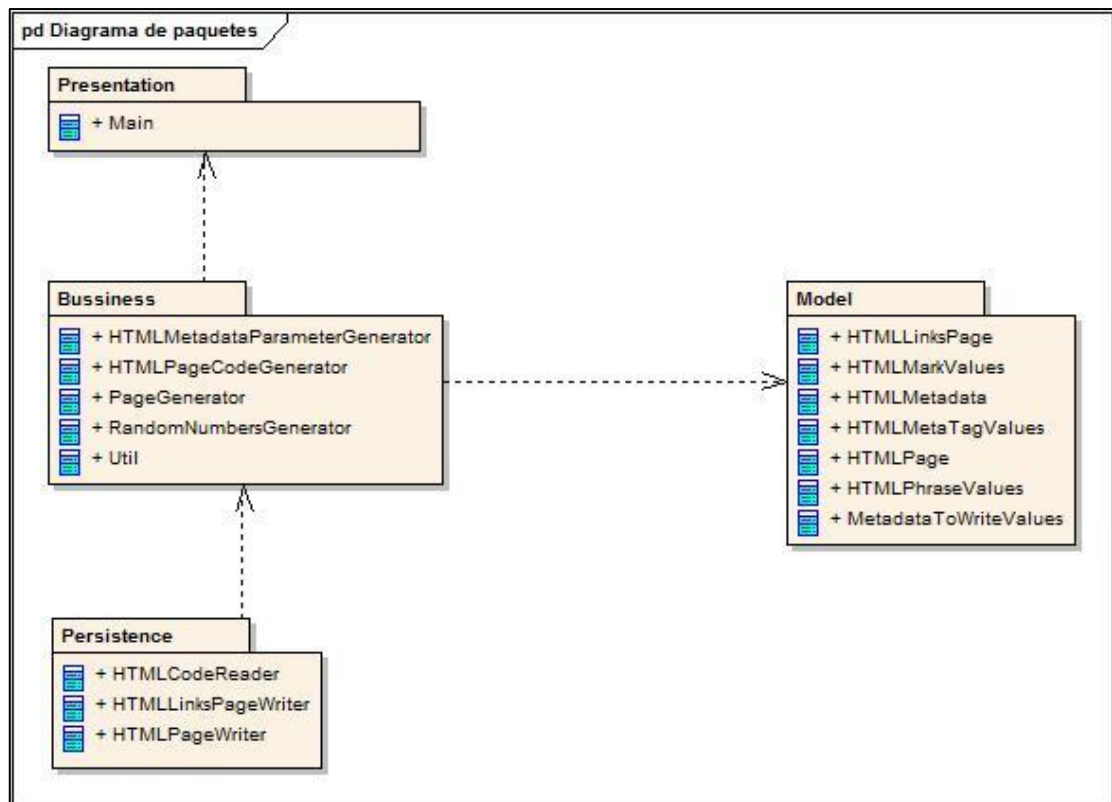


Diagrama 3.1 – Diagrama de paquetes del prototipo

Como se puede apreciar en el diagrama, para el diseño del prototipo se ha optado por una arquitectura por capas, las cuales son presentación, negocio y persistencia. Además, se ha creado un paquete específico para las clases del modelo.

En cuanto al paquete de presentación, al ser un prototipo meramente funcional a través de línea de comandos, tan sólo consta de una clase en la que se muestra una interfaz de usuario muy simple.

3.5.1.2 Diagramas de clases

3.5.1.2.1 Modelo

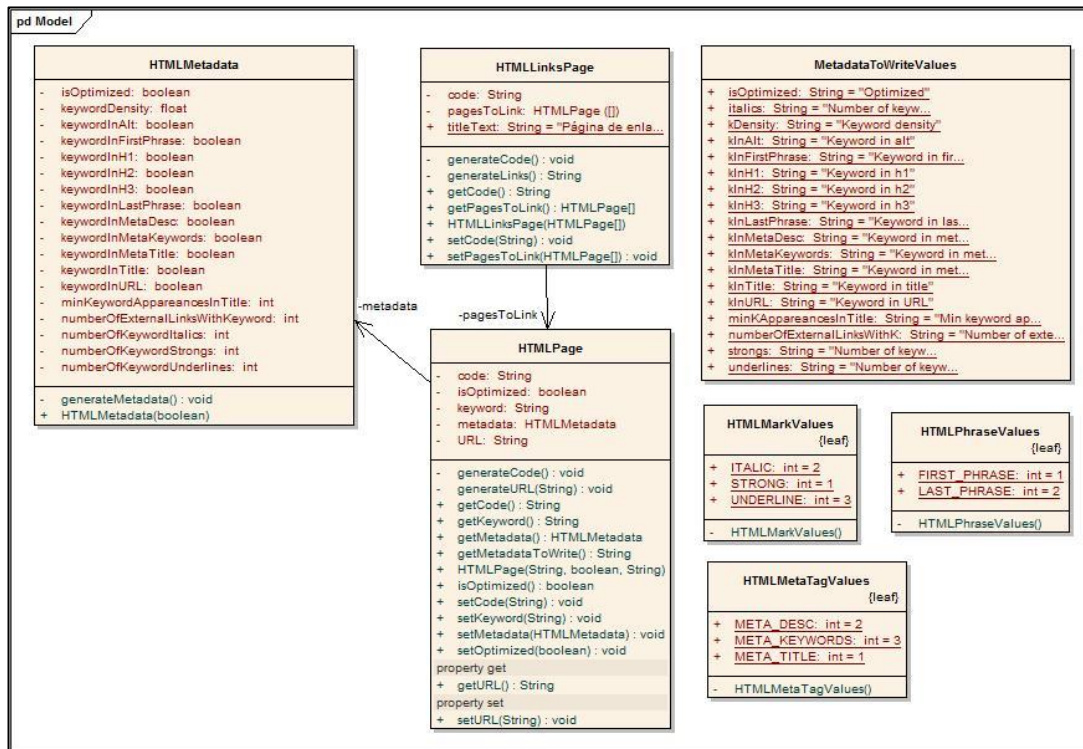


Diagrama 3.2 – Diagrama de clases del modelo del prototipo

Como se puede ver en el diagrama de clases del modelo del prototipo, hay tres clases fundamentales en el modelo:

- **HTMLMetadata:** en este tipo de objeto se almacenan diferentes metadatos de las páginas web generadas para que posteriormente se almacenarán en ficheros para que, una vez los buscadores nos devuelvan los resultados, poder tener algunos datos para obtener conclusiones. Como se puede ver en este objeto se almacenan atributos de la página asociada tales como si la palabra clave aparece en el título, la densidad de la palabra clave...
- **HTMLLinksPage:** En este tipo de objeto se almacenan las páginas que sirven de índice para que los buscadores accedan a las páginas web que forman parte de las muestras a estudiar. El atributo más importante de esta clase es el código de la misma, así como el vector donde se almacenan las páginas a enlazar. A partir de este objeto se generará un fichero HTML.
- **HTMLPage:** Este tipo de objeto sirve para almacenar las páginas a indexar por los buscadores y sobre las que se realizará el estudio de la investigación. Sus principales atributos son el código de la página y el objeto HTMLMetadata de la misma, donde se asocian los metadatos de la página web. A partir de estos objetos se generarán ficheros HTML que son los que indexarán los buscadores.

3.5.1.2.2 Capa de negocio

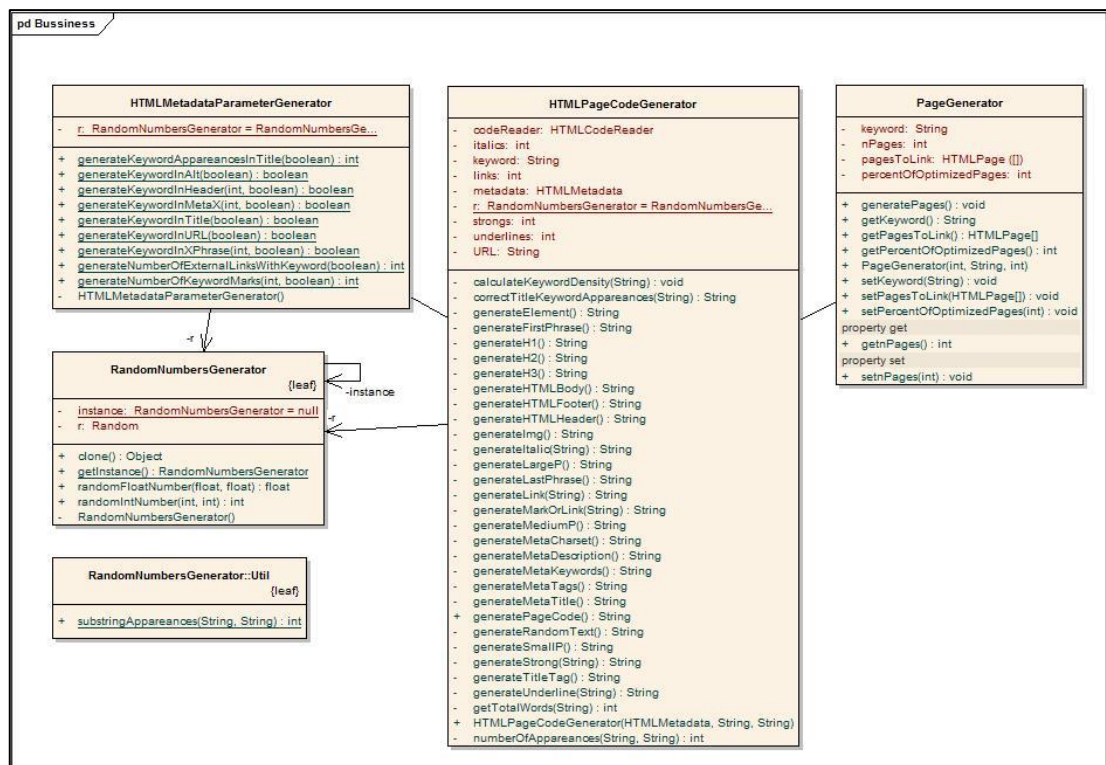


Diagrama 3.3 – Diagrama de la capa de negocio del prototipo

Las principales clases del modelo de negocio del prototipo son tres:

- **HTMLMetadataParameterGenerator**: Este tipo de objeto, gracias a algoritmos de generación aleatoria, es capaz de generar metadatos que tendrán las páginas web empleadas en las muestras.
- **HTMLPageCodeGenerator**: Este tipo de objeto emplea los metadatos generados por la anterior para generar, a partir de las muestras de textos asociadas a su palabra clave, diferentes elementos de la página web e insertarlos en el código de la misma.
- **PageGenerator**: Encargada de determinar el número de páginas a generar, así como sus URL y si serán optimizadas o no según los valores de entrada insertados por el usuario.

3.5.1.2.3 Capa de persistencia

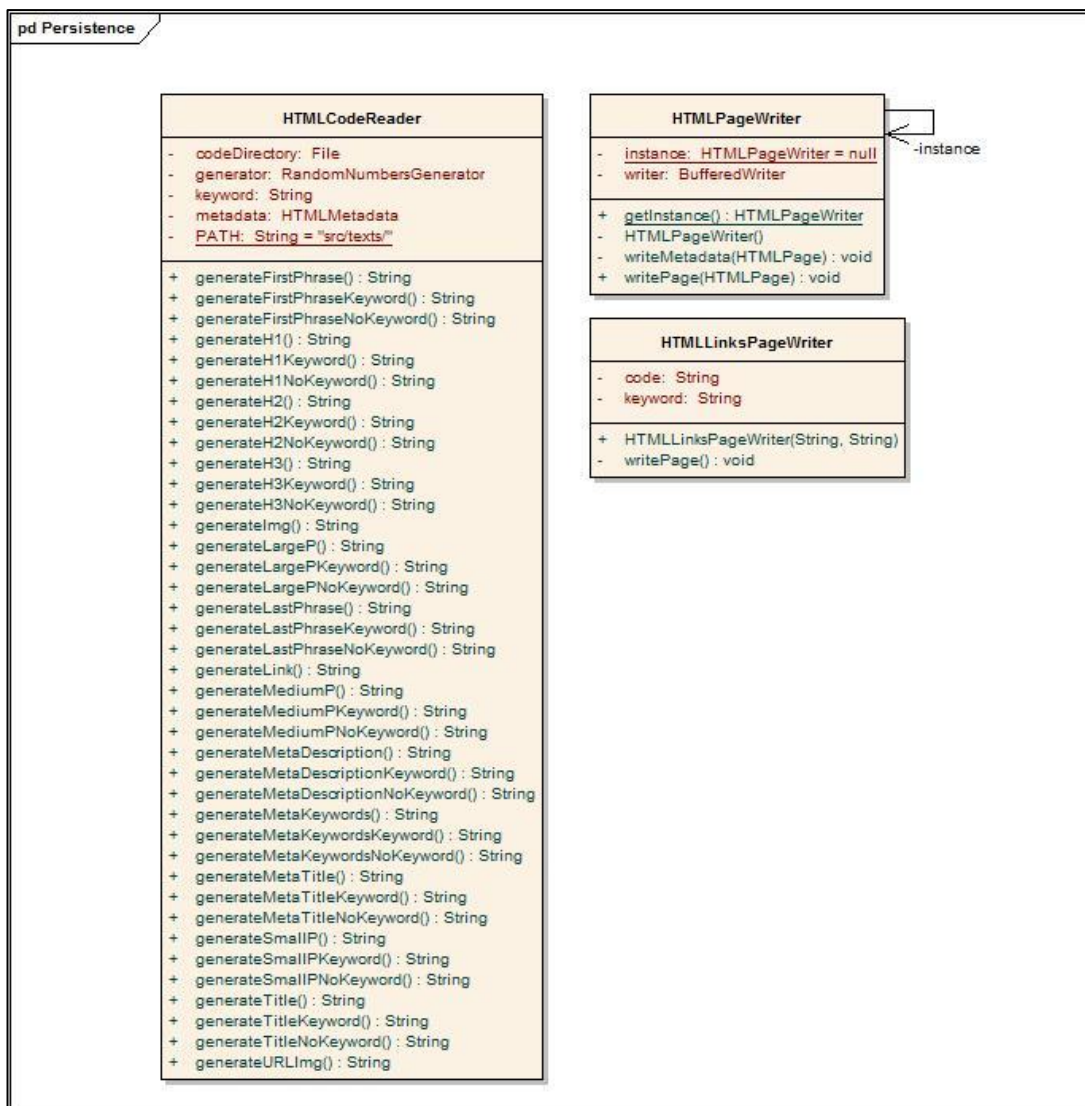


Diagrama 3.4 – Diagrama de la capa de persistencia del prototipo

Esta capa consta de tres clases, una de lectura y dos de escritura. Las dos clases más importantes son:

- **HTMLCodeReader:** Esta clase es la encargada de leer líneas de los ficheros de textos de muestra para las páginas web. Mediante la generación de un número aleatorio es capaz de leer diferentes líneas para diferentes elementos de un mismo tipo, como los títulos, de su fichero correspondiente.
- **HTMLPageWriter:** Esta clase se encarga, a partir de los datos del objeto de una página web, incluido su código, de generar los documentos HTML necesarios para las diferentes muestras de palabras clave a estudiar. La clase HTMLLinksPageWriter funciona de forma análoga a esta.

3.5.2 Valores de entrada del programa

El prototipo recibe tres valores de entrada para poder determinar cómo debe ser la muestra de páginas web a generar:

- La palabra clave a optimizar: esta palabra se sustituirá en todos los lugares indicados en los textos para generar las páginas web optimizadas.
- El número de páginas a generar: Para poder generar muestras de diferentes tamaños.
- El porcentaje de páginas web que deben ser optimizadas: A partir de este número se generarán del número total de páginas un porcentaje de ellas optimizadas para la palabra clave indicada y otras sin optimización.

3.5.3 Formato y organización de los textos de muestra

Las páginas web a generar emplearán textos de muestra para ello, que el prototipo se encargará de leer y organizar para crear páginas web que perfectamente podrían pasar por ejemplos reales redactados por una persona.

Estos textos de muestra estarán divididos en diferentes archivos, organizados por los diferentes elementos de los que consta un código HTML tradicional: encabezados, párrafos, imágenes, meta-etiquetas...

Además de esto, se dividen también en ficheros con muestras de texto optimizadas y con otras sin optimizar, de forma que, para la etiqueta de título, por ejemplo, habrá un fichero con diferentes muestras optimizadas y otro fichero con muestras sin optimizar.

En estos textos de muestra, en los lugares donde se debe insertar la palabra clave aparecerá el texto “=keyword=”. De esta forma se consigue que con una misma muestra de texto pudieran generarse diferentes textos finales para palabras clave distintas que encaje en ese lugar, o para estudiar por ejemplo el impacto de insertar una palabra clave en singular o en plural. Se ha buscado hacer el prototipo lo más reutilizable posible en este sentido.

Los ficheros en los cuales no aparecen elementos optimizados llevan la nomenclatura “[elemento].txt”, mientras que los que están optimizados siguen la nomenclatura “[elemento]key.txt”.

```
1 Variations: 10
2 Lorem ipsum =keyword=
3 Lorem ipsum =keyword= dolor
4 =keyword= Pellentesque habitant morbi
5 Pellentesque habitant =keyword=
6 Vestibulum =keyword= tortor quam, feugiat vitae
7 Vestibulum tortor quam, =keyword= feugiat
8 =keyword= Donec eu libero sit amet
9 Donec =keyword= eu sit amet
10 =keyword= Aenean ultricies mi =keyword= vitae est
11 Aenean ultricies =keyword= mi est
```

Ilustración 14 – Cómo son las muestras de texto de entrada optimizadas

```
1 Variations: 10
2 Lorem ipsum
3 Lorem ipsum dolor
4 Pellentesque habitant morbi
5 Pellentesque habitant
6 Vestibulum tortor quam, feugiat vitae
7 Vestibulum tortor quam, feugiat
8 Donec eu libero sit amet
9 Donec eu sit amet
10 Aenean ultricies mi vitae est
11 Aenean ultricies mi est
```

Ilustración 15 – Cómo son las muestras de texto de entrada sin optimizar

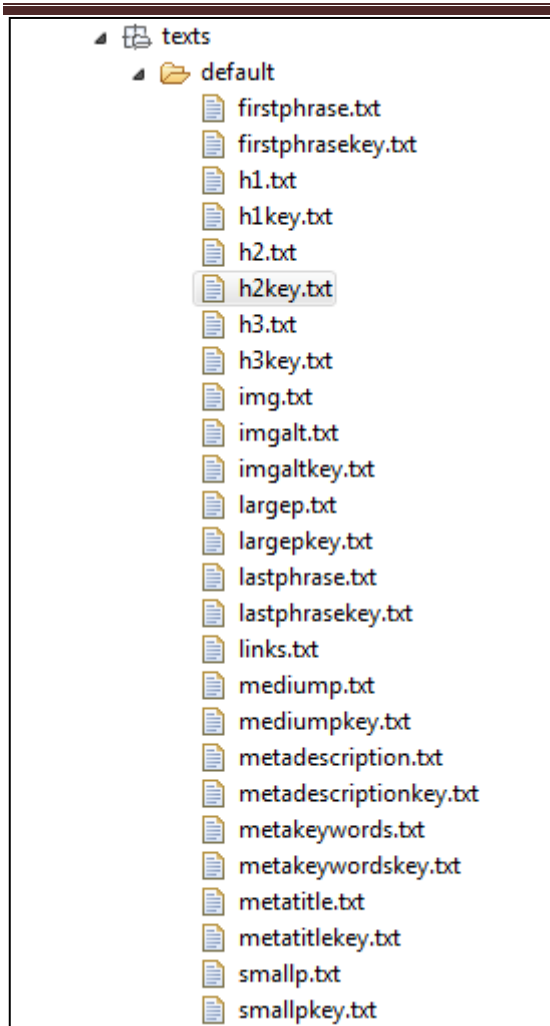


Ilustración 16 – Ficheros de entrada para cada palabra clave

Simplemente dando a la carpeta donde están los ficheros el nombre de la palabra clave para generar las muestras el prototipo ya es capaz de leer los textos de muestra de los ficheros mientras sigan la nomenclatura mencionada anteriormente.

3.5.4 Aleatoriedad en las páginas web generadas

Una de las partes más importantes de este proyecto es la forma en la que se generan las páginas y la aleatoriedad de los textos en las mismas, puesto que sin esto tendríamos gran cantidad de página, pero todas con el mismo texto, algo que es inútil para estudiar el comportamiento de los buscadores ante esas muestras de páginas.

Además, está el problema añadido del contenido duplicado que ya se comentó anteriormente en varios apartados, gracias al cual si tenemos dos o más páginas que son muy parecidas los buscadores tan solo tendrán en cuenta una de todas ellas.

Para simular la aleatoriedad y diversidad de textos que pueden encontrarse en Internet sobre un mismo tema se ha creado un algoritmo que es capaz, a partir de diferentes muestras de texto, generar páginas web con estructura y contenido diferente de forma automática.

Algunos de los trozos de código más relevantes de este algoritmo son los siguientes:

3.5.4.1 Generación de metadatos aleatoriamente

En esta clase se generan de forma aleatoria los metadatos que tendrán las páginas web a generar, por ejemplo:

- Si la palabra clave aparecerá en el título
- Si la palabra clave aparecerá en la meta-descripción
- Si la palabra clave aparece en la primera frase
- ...

```
public class HTMLMetadataParameterGenerator {
    private final static RandomNumbersGenerator r = RandomNumbersGenerator
        .getInstance();

    private HTMLMetadataParameterGenerator() {

    }

    public static boolean generateKeywordInXPhrase(int phrase, boolean optimize) {
        int randomValue = r.randomIntNumber(0, 10);
        switch (phrase) {
            case 1:
                if (optimize) {
                    if (randomValue > 2) {
                        return true;
                    } else {
                        return false;
                    }
                } else {
                    if (randomValue > 5) {
                        return true;
                    } else {
                        return false;
                    }
                }
            case 2:
                if (optimize) {
                    if (randomValue > 3) {
                        return true;
                    } else {
                        return false;
                    }
                } else {
                    if (randomValue > 6) {
                        return true;
                    } else {
                        return false;
                    }
                }
            default:
                if (optimize) {
                    if (randomValue > 5) {
                        return true;
                    } else {
                        return false;
                    }
                } else {
                    if (randomValue > 5) {
                        return true;
                    } else {
                        return false;
                    }
                }
        }
    }

    public static boolean generateKeywordInURL(boolean optimize) {
        if (optimize) {
            return true;
        } else {
            return false;
        }
    }
}
```

```
public static boolean generateKeywordInMetaX(int meta, boolean optimize) {
    int randomValue = r.randomIntNumber(0, 10);
    switch (meta) {
        case HTMLMetaTagValues.META_TITLE:
            return generateKeywordInTitle(optimize);
        case HTMLMetaTagValues.META_DESC:
            if (optimize) {
                if (randomValue > 2) {
                    return true;
                } else
                    return false;
            } else {
                if (randomValue > 4) {
                    return true;
                } else
                    return false;
            }
        case HTMLMetaTagValues.META_KEYWORDS:
            if (optimize) {
                if (randomValue > 3) {
                    return true;
                } else
                    return false;
            } else {
                if (randomValue > 6) {
                    return true;
                } else
                    return false;
            }
        default:
            if (optimize) {
                if (randomValue > 5) {
                    return true;
                } else
                    return false;
            } else {
                if (randomValue > 8) {
                    return true;
                } else
                    return false;
            }
    }
}

public static boolean generateKeywordInAlt(boolean optimize) {
    int randomValue = r.randomIntNumber(0, 10);
    if (optimize) {
        if (randomValue > 4) {
            return true;
        } else
            return false;
    } else {
        if (randomValue > 7) {
            return true;
        } else
            return false;
    }
}

public static int generateKeywordAppareancesInTitle(boolean optimize) {
    int randomValue = r.randomIntNumber(0, 10);
    if (optimize) {
        if (randomValue > 0 && randomValue < 3) {
            return 2;
        } else
            return 1;
    } else {
        if (randomValue > 8) {
            return 1;
        } else
            return 0;
    }
}

public static boolean generateKeywordInTitle(boolean optimize) {
```

```
int randomValue = r.randomIntNumber(0, 10);
if (optimize) {
    if (randomValue > 3) {
        return true;
    } else
        return false;
} else {
    if (randomValue > 7) {
        return true;
    } else
        return false;
}
}

public static int generateNumberOfKeywordMarks(int mark, boolean optimize) {
    switch (mark) {
        case 1:
            if (optimize) {
                return r.randomIntNumber(2, 5);
            } else {
                return r.randomIntNumber(0, 2);
            }

        case 2:
            if (optimize) {
                return r.randomIntNumber(1, 4);
            } else {
                return r.randomIntNumber(0, 1);
            }

        case 3:
            if (optimize) {
                return r.randomIntNumber(1, 3);
            } else {
                return r.randomIntNumber(0, 1);
            }

        default:
            if (optimize) {
                return r.randomIntNumber(1, 2);
            } else {
                return r.randomIntNumber(0, 1);
            }
    }
}

public static int generateNumberOfExternalLinksWithKeyword(boolean optimize) {
    int randomValue = r.randomIntNumber(0, 10);
    if (optimize) {
        if (randomValue > 7) {
            return 2;
        }
        if (randomValue > 2 && randomValue <= 7) {
            return 1;
        }
        return 0;
    } else {
        if (randomValue > 5) {
            return 1;
        } else
            return 0;
    }
}

public static boolean generateKeywordInHeader(int numberOfHeader,
        boolean optimize) {
    int randomValue = r.randomIntNumber(0, 10);
    switch (numberOfHeader) {
        case 1:
            if (optimize) {
                if (randomValue > 3) {
                    return true;
                } else
                    return false;
            } else {
                if (randomValue > 7) {
```

```
        return true;
    } else
        return false;
    }
case 2:
    if (optimize) {
        if (randomValue > 4) {
            return true;
        } else
            return false;
    } else {
        if (randomValue > 8) {
            return true;
        } else
            return false;
    }
case 3:
    if (optimize) {
        if (randomValue > 5) {
            return true;
        } else
            return false;
    } else {
        if (randomValue > 9) {
            return true;
        } else
            return false;
    }
default:
    if (optimize) {
        if (randomValue > 7) {
            return true;
        } else
            return false;
    } else
        return false;
}
}
```

Como se puede apreciar, en cada método se genera un metadato diferente, y en él se tiene en cuenta si la página seleccionada para generar los metadatos de esta tiene el atributo "optimize" con valor "true" o "false".

Estos metadatos serán posteriormente el punto de partida para generar los diferentes elementos de la página web, ya que en función de si un metadato es verdadero o falso o toma cierto valor se leerá de ficheros optimizados o no optimizados para seleccionar los textos de la página web.

3.5.4.2 Obtención de los textos en función de los metadatos

El siguiente fragmento de código ilustra cómo, en función de si un metadato indica que debe ser optimizado o no, se escoge leer de un fichero optimizado o de uno que no está optimizado:

```
public String generateTitle() throws IOException {
    File titleFile;
    if (metadata.isKeywordInTitle()) {
        titleFile = new File(codeDirectory.getPath() + "/titlekey.txt");
    } else
        titleFile = new File(codeDirectory.getPath() + "/title.txt");
    String title = readRandomLineOfFile(titleFile);
    title = replaceKeyword(title);
    return title;
}
```

De forma análoga a este método que genera el título optimizado o no optimizado se generan el resto de etiquetas, como la meta-descripción, las palabras clave, párrafos, encabezados...

Además de esto, se dispone de un método para determinar el elemento a generar en cada caso de forma aleatoria: de esta forma conseguimos generar documentos con diferente estructura. El método es el siguiente:

```
private String generateElement() throws IOException {
    int element = r.randomIntNumber(1, 15);
    switch (element) {
        case 1:
            return generateH2();

        case 2:
            return generateH3();

        case 3:
            return generateImg();

        case 4:
            return generateSmallP();

        case 5:
            return generateMediumP();

        case 6:
            return generateLargeP();

        default:
            element = r.randomIntNumber(1, 10);
            if (element >= 1 && element < 3) {
                return generateLargeP();
            } else if (element >= 3 && element < 7) {
                return generateMediumP();
            } else
                return generateSmallP();
    }
}
```

También se dispone de un método para poder generar páginas con diferentes números de elementos, dentro de un mínimo y un máximo establecidos que se asemejen al número de elementos medio que tiene una página web real:

```
private String generateRandomText() throws IOException {
    String randomText = "";
    int numberOfElements = r.randomIntNumber(10, 15);
    for (int i = 0; i < numberOfElements; i++) {
        randomText += generateElement();
    }
    return randomText;
}
```

3.5.5 Resultados del prototipo

A partir de los valores de entrada mencionados anteriormente, de los textos de muestra y de los algoritmos explicados en el apartado anterior, el prototipo es capaz de generar tantas páginas web optimizadas para una palabra clave como se le indique en segundos, tal y como se puede ver en la siguiente imagen:





























	20	13/01/2013 10:11	Chrome HTML Do...	8 KB
	21	13/01/2013 10:11	Chrome HTML Do...	8 KB
	22	13/01/2013 10:11	Chrome HTML Do...	8 KB
	23	13/01/2013 10:11	Chrome HTML Do...	8 KB
	24	13/01/2013 10:11	Chrome HTML Do...	8 KB
	25	13/01/2013 10:11	Chrome HTML Do...	8 KB
	26	13/01/2013 10:11	Chrome HTML Do...	8 KB
	27	13/01/2013 10:11	Chrome HTML Do...	4 KB
	28	13/01/2013 10:11	Chrome HTML Do...	4 KB
	29	13/01/2013 10:11	Chrome HTML Do...	8 KB
	30	13/01/2013 10:11	Chrome HTML Do...	8 KB
	31	13/01/2013 10:11	Chrome HTML Do...	8 KB
	32	13/01/2013 10:11	Chrome HTML Do...	4 KB
	33	13/01/2013 10:11	Chrome HTML Do...	8 KB
	34	13/01/2013 10:11	Chrome HTML Do...	8 KB
	35	13/01/2013 10:11	Chrome HTML Do...	8 KB
	36	13/01/2013 10:11	Chrome HTML Do...	8 KB
	37	13/01/2013 10:11	Chrome HTML Do...	4 KB
	38	13/01/2013 10:11	Chrome HTML Do...	4 KB
	39	13/01/2013 10:11	Chrome HTML Do...	8 KB
	links-noticias	13/01/2013 10:11	Chrome HTML Do...	4 KB
	noticias-0	13/01/2013 10:11	Chrome HTML Do...	8 KB
	noticias-1	13/01/2013 10:11	Chrome HTML Do...	8 KB
	noticias-2	13/01/2013 10:11	Chrome HTML Do...	8 KB
	noticias-3	13/01/2013 10:11	Chrome HTML Do...	8 KB
	noticias-4	13/01/2013 10:11	Chrome HTML Do...	8 KB
	noticias-5	13/01/2013 10:11	Chrome HTML Do...	8 KB
	noticias-6	13/01/2013 10:11	Chrome HTML Do...	8 KB

Ilustración 17 – Páginas web generadas por el prototipo para la palabra clave “noticias”

Como se puede ver, hay tres tipos de páginas generadas:

- El nombre de archivo es un número: estas páginas web son las que no han sido optimizadas. Como puede verse en la imagen, la palabra noticias aparece relativamente poco, y apenas se puede ver en encabezados. Tampoco hay imágenes

con la palabra clave en la página, aunque podría haberlas.



Ilustración 18 – Apariencia de una página web generada no optimizada

- El nombre de archivos tiene la estructura **[keyword]-[número]**: estas páginas web son las que han sido optimizadas. Como puede verse en la siguiente imagen, en los encabezados ya aparece más la palabra clave, así como en negritas, cursivas, etc. También puede verse una imagen cuyo texto alternativo incluye la palabra clave (la imagen no aparece porque no existe el archivo, pero eso es irrelevante para los buscadores), así como un enlace externo.



Ilustración 19 – Apariencia de una página web generada optimizada

- **Links-[keyword]**: es la página índice de la que hablamos en apartados anteriores.

Como puede verse en las imágenes, las páginas web generadas no son visualmente muy estéticas, pero eso es irrelevante para esta investigación, ya que lo único que las diferenciaría de páginas web reales sería un archivo .css para darles formato, el resto sería todo igual. Esto no tiene relevancia porque, a pesar de que se cree que los buscadores actualmente sí son capaces de visualizar páginas web como lo haría un ser humano, si todas las páginas de la muestra tienen una apariencia similar todas serán puntuadas en este aspecto de la misma forma, y es un factor que no estudiaremos.

Además de esto, el prototipo genera un archivo .properties por cada página web indicando algunos de sus metadatos, aunque no siempre es exacto. Un ejemplo de este archivo generado sería el siguiente:

```
Optimized:No
Keyword in title:No
Keyword in alt:No
Keyword in first phrase:Yes
Keyword in last phrase:No
Keyword in h1:No
Keyword in h2:Yes
Keyword in h3:No
Keyword in meta description:Yes
Keyword in meta keywords:Yes
Keyword in meta title:Yes
Keyword in URL:No
Min keyword appareances in title:0
Number of external links with keyword:1
Number of keyword italics:0
Number of keyword strongs:0
Number of keyword underlines:0
Keyword density:0.022988506
```

- “*Optimized*” indica si la página web se considera que está optimizada o no por el algoritmo. Aunque el algoritmo tiene cierto componente aleatorio, si una página web se indica que no está optimizada esta tiene muchas menos probabilidades de contener elementos optimizados en ella, aunque en casos extremos puede ocurrir que una no optimizada en principio resulte más optimizada que una que debería estarlo. Esto es algo que se contempla dentro del mundo real donde sin querer se pueden hacer páginas web optimizadas o viceversa.
- El resto de parámetros no son siempre exactos, puede haber ocasiones en que a veces no se correspondan con la realidad, salvo los siguientes que son siempre exactos: *keyword in meta title*, *keyword in meta description*, *keyword in meta keywords*, *keyword density*, *number of keyword italics*, *strongs and underlines*. Por lo tanto se estudiará cada página web en particular cuando se tomen resultados para hacer un buen análisis de los mismos.
- El parámetro *keyword in meta title* siempre que esté a “Yes” se cumplirá, pero si está en “No”, el factor suerte influye en que la palabra clave aparezca en el título o no.

3.6 Elección y creación de muestras de palabras clave a estudiar

Una vez creado nuestro generador de páginas web hay que elegir las palabras clave que se van a estudiar a utilizar en esta investigación.

En concreto se elegirán unas palabras clave para estudiar las posiciones que devuelven los buscadores en relación a estas y las páginas web que hemos creado.

Para la elección de palabras clave tendremos en cuenta las siguientes normas:

- Estudiar al menos 4 palabras clave para poder afirmar con certeza que las conclusiones que sacamos son globales y no fruto de casualidades.
- Elegir palabras y textos en, al menos, dos idiomas diferentes. Esto no quiere decir que haya que estudiar la misma palabra clave en dos idiomas, sino que conviene que, en total, se estudien varios idiomas.
- Elegir algunas palabras clave de temáticas variadas para que no se solapen en una búsqueda páginas de dos palabras clave diferentes.
- Elegir otras palabras clave similares para estudiar el solapamiento e intrusión de páginas web en teoría posicionadas para una palabra clave en los resultados de búsqueda de otra *keyword*.
- Evitar que las palabras clave posean tildes, por si los buscadores discrepan de la palabra con tilde o sin tilde, cosa que aún se desconoce.
- Evitar palabras clave de temáticas muy genéricas que puedan solaparse con todas las demás, como por ejemplo “*noticia*” o “*artículo*”.
- Las palabras clave deben estar en singular y, si puede ser, en su género masculino.

3.6.1 Palabras clave elegidas

De esta forma, la lista de palabras clave sobre las que se va a experimentar es la siguiente:

- Dieta.
- Adelgazar: se solapa con dieta, lo cual es importante para estudiar relaciones entre palabras clave de una misma temática.
- Computador.
- Car.

3.6.2 Creación de los ficheros de muestra

Para cada muestra habrá que crear todos los archivos necesarios para que el prototipo sea capaz de generar todo tipo de elementos en las páginas web, y variando archivos con muestras optimizadas y otras que no lo estén. El conjunto de archivos que debe poseer cada palabra clave para la generación de su muestra es el siguiente:

Nombre	Fecha de modifica...	Tipo	Tamaño
firstphrase	04/10/2012 10:28	Documento de tex...	1 KB
firstphrasekey	04/10/2012 10:28	Documento de tex...	1 KB
h1	04/10/2012 10:25	Documento de tex...	1 KB
h1key	08/10/2012 20:02	Documento de tex...	1 KB
h2	04/10/2012 10:28	Documento de tex...	1 KB
h2key	04/10/2012 10:28	Documento de tex...	1 KB
h3	04/10/2012 10:28	Documento de tex...	1 KB
h3key	04/10/2012 10:28	Documento de tex...	1 KB
img	08/10/2012 12:24	Documento de tex...	1 KB
imgalt	06/10/2012 20:40	Documento de tex...	1 KB
imgaltkey	06/10/2012 20:40	Documento de tex...	1 KB
largep	04/10/2012 10:30	Documento de tex...	2 KB
largepkey	06/10/2012 18:28	Documento de tex...	2 KB
lastphrase	04/10/2012 10:30	Documento de tex...	1 KB
lastphrasekey	04/10/2012 10:30	Documento de tex...	1 KB
links	06/10/2012 19:59	Documento de tex...	1 KB
mediump	04/10/2012 10:36	Documento de tex...	1 KB
mediumpkey	06/10/2012 18:30	Documento de tex...	1 KB
metadescription	04/10/2012 15:48	Documento de tex...	1 KB
metadescriptionkey	04/10/2012 15:48	Documento de tex...	1 KB
metakeywords	04/10/2012 15:50	Documento de tex...	1 KB
metakeywordskey	04/10/2012 15:50	Documento de tex...	1 KB
metatitle	04/10/2012 15:46	Documento de tex...	1 KB
metatitlekey	04/10/2012 15:47	Documento de tex...	1 KB
smallp	04/10/2012 10:36	Documento de tex...	1 KB
smallpkey	04/10/2012 10:36	Documento de tex...	1 KB
title	11/10/2012 13:10	Documento de tex...	1 KB
titlekey	11/10/2012 11:33	Documento de tex...	1 KB

Ilustración 20 –Archivos necesarios para generar muestras de cada palabra clave

Para los diferentes textos a incluir en cada archivo, estos se extraerán de páginas web reales, a poder ser que esté verificada la autenticidad de sus contenidos y que estén bien redactados, como puede ser por ejemplo Wikipedia, páginas de universidades o instituciones públicas.

3.6.3 Ejemplo de ficheros de muestra

A continuación se muestra un ejemplo de cómo debe ser cada uno de los archivos de textos de muestra, en este caso para la palabra clave “dieta”, ordenados alfabéticamente.

3.6.3.1 *firstphrase.txt*

Variations: 5

La piña, además de ser una fruta tropical y presumir de un estupendo sabor, contiene propiedades muy saludables para el organismo. Entra en este artículo enviado por una usuaria y disfruta de sus ventajas nutricionales.

El arroz y la fruta pueden ser tus mejores aliados en esa dura batalla que te espera a la hora de perder peso.

Esta fruta afrodisíaca ayuda a la desintoxicación del organismo y contribuye a mejorar la calidad de la piel.

La naranja es una de las frutas más depurativa y rica en vitaminas.

La sandía está considerada como una de las frutas más deliciosas y refrescantes.

3.6.3.2 *firstphrasekey.txt*

Variations: 5

La =keyword= de la remolacha te permitirá perder peso en muy poco tiempo, siempre y cuando la respetes a rajatabla.

Ahora con la llegada del verano, esta =keyword= te ayudará a bajar esos kilos de más y te aportará muchas de las propiedades que posee.

El alimento base en esta =keyword= es la zanahoria una hortaliza rica en fibra ideal para bajar de peso, entre 3 a 4 kilos.

¿Quieres perder hasta 3 kilos en 15 días mientras aprovechas para depurar tu cuerpo de toxinas? Te contamos más sobre la =keyword= de los cítricos.

Creada en Japón por Hitoshi Watanabe para ayudarlo a perder peso, esta =keyword= permite perder hasta 3-5 kilos en 2 semanas.

3.6.3.3 *h1.txt*

Variations: 10

Choque de la piña

Perder peso fácilmente

Bajar esos kilos de más

Normas para adelgazar

Normas básicas de adelgazamiento

Deja de estar gordo

¿Cómo adelgazar sin riesgo para la salud?

Perder esos kilos de más de cara al verano

Adelgaza haciendo deporte y cuidando la alimentación

Alimentos recomendados para perder peso

3.6.3.4 *h1key.txt*

Variations: 10

=keyword= (alimentación)

Historia de las =keyword=s humanas

Tipos de =keyword=s contemporáneas

=keyword= vegetariana

=keyword= omnívora

=keyword= carnívora

=keyword=s basales o =keyword=s básicas

=keyword=s terapéuticas

La =keyword= saludable y equilibrada

Animales por =keyword=

3.6.3.5 h2.txt

```
Variations: 10
Bajando de peso
Adelgazar es fácil
Consejos para mantener el peso
Alimenta tu cuerpo de forma saludable
¿Cómo puedes mejorar tu alimentación?
¿Es posible adelgazar comiendo hamburguesas?
Evitar los productos con muchas grasas
Come frutas y verduras
Evitar las grasas
Pierde kilos de más y elimina líquidos con nuestros consejos
```

3.6.3.6 h2key.txt

```
Variations: 10
=keyword=s Hipocalóricas para bajar peso
=keyword=s abiertas
=keyword=s por Intercambios
=keyword=s cerradas
=keyword= para pacientes con Dislipemia o Hipercolesterolemia
=keyword= para pacientes con Triglicéridos elevados
=keyword= pobre en Purinas (para pacientes con hiperuricemia o gota)
=keyword=s bajas en sal
=keyword= sin gluten
Preparación y =keyword=s en las exploraciones
```

3.6.3.7 h3.txt

```
Variations: 10
Choque de la piña
Perder peso fácilmente
Bajar esos kilos de más
Normas para adelgazar
Normas básicas de adelgazamiento
Deja de estar gordo
¿Cómo adelgazar sin riesgo para la salud?
Perder esos kilos de más de cara al verano
Adelgaza haciendo deporte y cuidando la alimentación
Alimentos recomendados para perder peso
```

3.6.3.8 h3key.txt

```
Variations: 10
=keyword= (alimentación)
Historia de las =keyword=s humanas
Tipos de =keyword=s contemporáneas
=keyword= vegetariana
=keyword= omnívora
=keyword= carnívora
=keyword=s basales o =keyword=s básicas
=keyword= para pacientes con Dislipemia o Hipercolesterolemia
=keyword= para pacientes con Triglicéridos elevados
=keyword= pobre en Purinas (para pacientes con hiperuricemia o gota)
```

3.6.3.9 img.txt

```
Variations: 3
img1.jpg
img2.jpg
img3.jpg
```

3.6.3.10 *imgalt.txt*

```
Variations: 3



```

3.6.3.11 *imgaltkey.txt*

```
Variations: 3



```

3.6.3.12 *largep.txt*

```
Variations: 5
```

Puede incrementar ampliamente la calidad de tus entrenamientos si mejoras la manera en que respiras. Mucha gente utiliza menos del 30% de su capacidad pulmonar. Aprender a respirar bien es vital para la salud de tu corazón, del metabolismo y muchas otras funciones. Sería una buena apuesta aprender algunos métodos respiratorios, ya que esto puede mejorar tu estado físico y tu salud, tu desempeño en los deportes o actividades físicas y en general en todo en tu vida.

El ejercicio con pesas es fantástico para incrementar y mejorar tu estado físico. Estimula la circulación, mejora la oxigenación de los músculos y tejidos, y acelera el metabolismo. Es excelente para el corazón y también para quemar grasas. Si puedes entrenarte al menos 3 veces a la semana durante 45 - 60 minutos, te aseguras grandes beneficios.

Evita los estimulantes como el alcohol, los cigarrillos o cualquier droga, cualquiera de estas cosas le hacen un tremendo daño a tu salud y reducen notablemente la producción natural de energía del cuerpo. Comer a la misma hora es importante ya que asegura un funcionamiento óptimo de tu cuerpo, incluyendo el funcionamiento hormonal y celular. Para conseguir el mejor funcionamiento de tu cuerpo come cada tres horas, comenzando bien temprano por las mañanas, para un total de 5 o 6 comidas.

Reduce (en lo posible elimina) el sodio, los azúcares refinados, y los carbohidratos refinados y procesados. Estos productos son responsables de inflamación, niveles de insulina altos y presión sanguínea alta entre otras cosas. Van en detrimento de tu salud cardiovascular y si los evitas notarás mejoras asombrosas. Estos sólo son algunos de los simples consejos que puedes seguir para obtener mejores resultados y mejorar tu estado físico y forma. Pasos pequeños para grandes resultados..

Durante ese tiempo, el sistema cardiovascular deberá ser capaz de mantener un aporte adecuado de oxígeno y nutrientes, tanto a la musculatura en activo como al resto de los órganos de nuestro cuerpo. Este aspecto de la actividad física es el que parece proporcionar la mayoría de los beneficios para la salud derivados de la práctica de ejercicio, y por ello constituye la piedra angular de cualquier programa de acondicionamiento físico. A través del entrenamiento, el corazón, los pulmones, las arterias y otros órganos de nuestro cuerpo se adaptan para trabajar conjuntamente de forma más eficiente en respuesta al estrés que supone la actividad física.

3.6.3.13 *largepkey.txt*

Variations: 5
TengoAgujetas es una página web dedicada a todos aquellos amantes del deporte o aquellos que quieran iniciarse. Además también tratamos el complicado tema de las =keyword=s desde diferentes puntos de vista: =keyword=s para adelgazar, para ganar volumen en el gimnasio... por cierto, ¡también nos encanta el gimnasio y publicamos rutinas y ejercicios semanalmente!
Si estás aquí seguramente es porque estás un poco desorientado en el tema de cómo ganar masa muscular en el gimnasio o, simplemente, en busca de mejorar tus hábitos y tu =keyword= de cara a obtener un mayor volumen para tus músculos, y por eso vamos a darte cinco consejos básicos que debes cumplir cada día para obtener importantes ganancias de masa muscular en tu cuerpo.
Si ya estás un poco metido en la materia quizá estos consejos te resulten repetitivos, entonces te pedimos disculpas, pero entiende que hay personas a las que hasta ahora nadie les ha informado de las cosas que deben tener en cuenta a la hora de seguir una =keyword= de volumen muscular.
Como ya te comentamos, hay unos alimentos con mejores propiedades que otros para según que cosas. Concretamente, si quieres ganar volumen, tu =keyword= debe centrarse en alimentos ricos en carbohidratos complejos, proteínas y grasas saludables. Debes comer de todo, pero sobre todo estos alimentos.
¿Estás buscando una =keyword= de volumen? Aquí te ofrecemos un ejemplo de =keyword= con la que conseguirás ingerir 3.000 kcal diarias. Recuerda que la cantidad de calorías que necesitas viene determinada por tu peso, tu actividad física y tu metabolismo. Esta =keyword= está pensada para una persona que pesa unos 75 kg de media y quiere ganar volumen muscular, por lo que si es tu caso toma nota:

3.6.3.14 *lastphrase.txt*

Variations: 5
Al principio te costará nadar si no sabes, pero con un par de cursillos te moverás como pez en el agua, nunca mejor dicho.
Mientras que en otros deportes como la bici o correr el ejercicio siempre es el mismo con pequeñas variaciones en ritmo e intensidad, nadando podemos variar no solo eso, sino el ejercicio completo.
El brócoli es un vegetal que pertenece a la familia de las Crucíferas, como la col o el repollo, y tiene un sabor amargo que a muchas personas no les agrada en absoluto. En cuanto a las vitaminas, el brócoli contiene gran cantidad de vitaminas A y C, esenciales para nuestra actividad diaria.
Los ya comentados antioxidantes también ayudan a prevenir enfermedades cancerígenas según se ha concluido en numerosas investigaciones, pero es que además el brócoli tiene grandes cantidades de antiestrógenos naturales.

3.6.3.15 *lastphrasekey.txt*

Variations: 5
Por eso la mayoría de =keyword=s de definición, al igual que esta, son bastante estrictas y conviene variarlas cada poco tiempo si no se quiere caer en una rutina que te llevará a dejarlo.
Allá vamos con esta =keyword= de definición:
No olvides romper la =keyword= una vez por semana haciendo una "comida trampa".
Si te ha gustado y quieres mantenerte al día con los últimos consejos deportivos, artículos sobre nutrición y =keyword=s puedes suscribirte por correo electrónico a nuestras publicaciones:
3 razones para incluir el té verde en tu =keyword= para adelgazar

3.6.3.16 *links.txt*

Variations: 3
<http://www.dietas.net>
<http://www.hagodieta.com>
<http://www.masquedietas.es>

3.6.3.17 *mediump.txt*

Variations: 5

Cuando uno habla de "alimentación y deporte", lo primero que se le ocurre es pensar en comer lo mejor posible para así desarrollar una actividad extra sin sufrir agotamiento físico. Sin embargo, no es tan sencillo: aunque comamos un día bien, seguiremos sin ser capaces de correr los 100 metros lisos sin caer en estado de fatiga.

Es más, la alimentación y el deporte deben recorrer juntos un largo camino para llegar al podio o simplemente para sentirse bien. Asimismo, hay que tener en cuenta que, además de recorrer un camino juntos, la alimentación y el deporte deben estar relacionados.

Es decir, no todos los deportes llevan asociado el mismo tipo de alimentación. Se deben hacer diferencias según una serie de características como tipo de acto deportivo, aeróbico o no, características del deportista, etc.

La alimentación del deportista no sólo debe nutrir las células del organismo para que éste se desarrolle y mantenga, sino que, además, debe cubrir el gasto derivado del esfuerzo extra.

También, como en otros aspectos de la vida, la alimentación del deportista se ve rodeada de ciertos matices que, si bien benefician en algunos casos, son muy perjudiciales en otros.

3.6.3.18 *mediumpkey.txt*

Variations: 5

¿Sigues una =keyword=? Todas lo hemos hecho, hemos decidido alguna vez empezar una =keyword=, la hemos abandonado, la hemos vuelto a seguir...

Para conseguir hacer una =keyword= de manera eficaz, no hay ningún misterio: evidentemente hace falta cierta motivación, pero también saber elegir la =keyword= adecuada, adaptada a nuestra edad, a nuestro estilo de vida, ¡y a la cantidad de kilos que hay que perder!

¿Tienes ganas de seguir una =keyword= que te haga perder kilos de manera rápida?

¡Busca en nuestras propuestas de =keyword=s exprés!

Las =keyword=s rápidas siempre han de estar vigiladas por un médico y debes combinarlas con ejercicio. Existen multitud de =keyword=s como la =keyword= del limón que te permite eliminar toxinas, eliminar los michelines superfluos y aumentar la luminosidad.

La =keyword= express te permitirá perder peso rápidamente en tres fases. Pero recuerda que las =keyword=s "de choque" no pueden seguirse durante mucho tiempo (un período máximo de 3 ó 4 semana), para que el organismo no se debilite ni tenga carencias.

3.6.3.19 *metadescription.txt*

Variations: 4

Alimentación es la ingestión de alimento por parte de los organismos para proveerse de sus necesidades alimenticias, fundamentalmente para conseguir energía y desarrollarse. El Ministerio de Agricultura, Alimentación y Medio Ambiente del Gobierno de España presenta el portal 'Alimentación.es', un escaparate de los alimentos de todos los sectores y todas las Comunidades Autónomas, mostrando el gran patrimonio agrario, alimentario y gastronómico nacional.

Los contenidos en vitaminas, tablas de calorías, sus efectos en la salud: colesterol, grasas, antioxidantes, ácido úrico, etc. serán decisivos en una nutrición rica en nutrientes.

Las guías alimentarias, como la pirámide de la alimentación saludable, sitúan al chocolate en su vértice, es decir, en el grupo de alimentos de consumo mínimo.

3.6.3.20 *metadescriptionkey.txt*

Variations: 4

=keyword=s para perder peso facilmente comiendo sano y sin pasar hambre.

Miles de =keyword=s para adelgazar y bajar de peso, así como también videos de ejercicios para perder peso, recetas y herramientas.

Los tipos de =keyword=s para adelgazar más populares explicadas de forma sencilla.

Descarga nuestro ebook gratis y pierde hasta 4,5 kilos en 2 semanas.

=keyword=s para adelgazar. En el "mercado" existen muchos métodos o formas de adelgazar que aquí te presentamos.

3.6.3.21 *metakeywords.txt*

```
Variations: 4
alimentacion, nutricion, adelgazar
perder peso, alimentacion, azúcares, grasas
perder grasa, adelgazar, abdominales
gluteos, pechos, definir
```

3.6.3.22 *metakeywordskey.txt*

```
Variations: 4
=keyword=, adelgazar, perder peso
adelgazar, =keyword= para adelgazar, pérdida de peso
=keyword=, azucar, grasa
abdominales, =keyword=
```

3.6.3.23 *metatitle.txt*

```
Variations: 10
Alimentación y nutrición
¿Qué es la alimentación?
Cómo llevar una alimentación saludable
¿Por qué es importante comer bien?
Somos lo que comemos
Nutrición y alimentación
Aliméntate bien
Cómo mantenerte en forma con tu alimentación
Alimentación equilibrada
Evita las grasas para adelgazar
```

3.6.3.24 *metatitlekey.txt*

```
Variations: 10
=keyword= para adelgazar
=keyword= para ganar volumen muscular
=keyword= de la manzana
Por qué incorporar el arroz en tu =keyword=
Hacer =keyword= es fundamental
Haz =keyword= y siéntete mejor
=keyword=s y nutrición
La pasta en nuestra =keyword=
=keyword=: CH, grasas y proteínas
4 formas de meter proteínas a tu =keyword=
```

3.6.3.25 *smallp.txt*

```
Variations: 5
No hay que confundir alimentación con nutrición, ya que esta última se da a nivel celular y la primera es la acción de ingerir un alimento. La nutrición puede ser autótrofa o heterótrofa.
Las plantas son organismos autótrofos. Son capaces de sintetizar compuestos orgánicos a partir de sales minerales y de la energía solar a través de la función clorofílica. Como otros animales, el hombre depende de su medio ambiente para asegurar sus necesidades fundamentales de alimento.
Los contenidos en vitaminas, tablas de calorías, sus efectos en la salud.
Te seleccionamos una lista de alimentos que por muchos especialistas en estos temas son considerados alimentos afrodisiacos....
```

3.6.3.26 *smallpkey.txt*

```
Variations: 5
=keyword= basada en la ingestión de 5 comidas al día, para que el cuerpo obtenga todas las proteínas necesarias para el organismo.
Esta =keyword= te indica las pautas para conseguir masa muscular combinándola con un entrenamiento diario y una nutrición específica para obtener los resultados deseados.
Los deportistas necesitan una =keyword= rica en carbohidratos ya que han de mantener la disposición de glucógeno muscular durante el ejercicio.
Los deportistas suelen tener un gran gasto de energía por todo el esfuerzo físico realizado. A continuación te ofrecernos una =keyword= de alto nivel de glucógeno para afrontar las largas sesiones de entrenamiento y competiciones.
Esta =keyword= basada en alimentos proteínicos y equilibrados te ayudará a ganar esos kilos que necesitas ingiriendo 6 comidas diarias.
```

3.6.3.27 *title.txt*

```
Variations: 10
Alimentación y nutrición
¿Qué es la alimentación?
Cómo llevar una alimentación saludable
¿Por qué es importante comer bien?
Somos lo que comemos
Nutrición y alimentación
Aliméntate bien
Cómo mantenerte en forma con tu alimentación
Alimentación equilibrada
Evita las grasas para adelgazar
```

3.6.3.28 *titlekey.txt*

```
Variations: 10
=keyword= para adelgazar
=keyword= para ganar volumen muscular
=keyword= de la manzana
Por qué incorporar el arroz en tu =keyword=
Hacer =keyword= es fundamental
Haz =keyword= y siéntete mejor
=keyword=s y nutrición
La pasta en nuestra =keyword=
=keyword=: CH, grasas y proteínas
4 formas de meter proteínas a tu =keyword=
```

Una vez tenemos todos los textos de muestra creados procedemos a crear las diferentes páginas web para cada palabra clave. Para ello, desde el eclipse abrimos el proyecto y configuramos una serie de parámetros de ejecución del main:

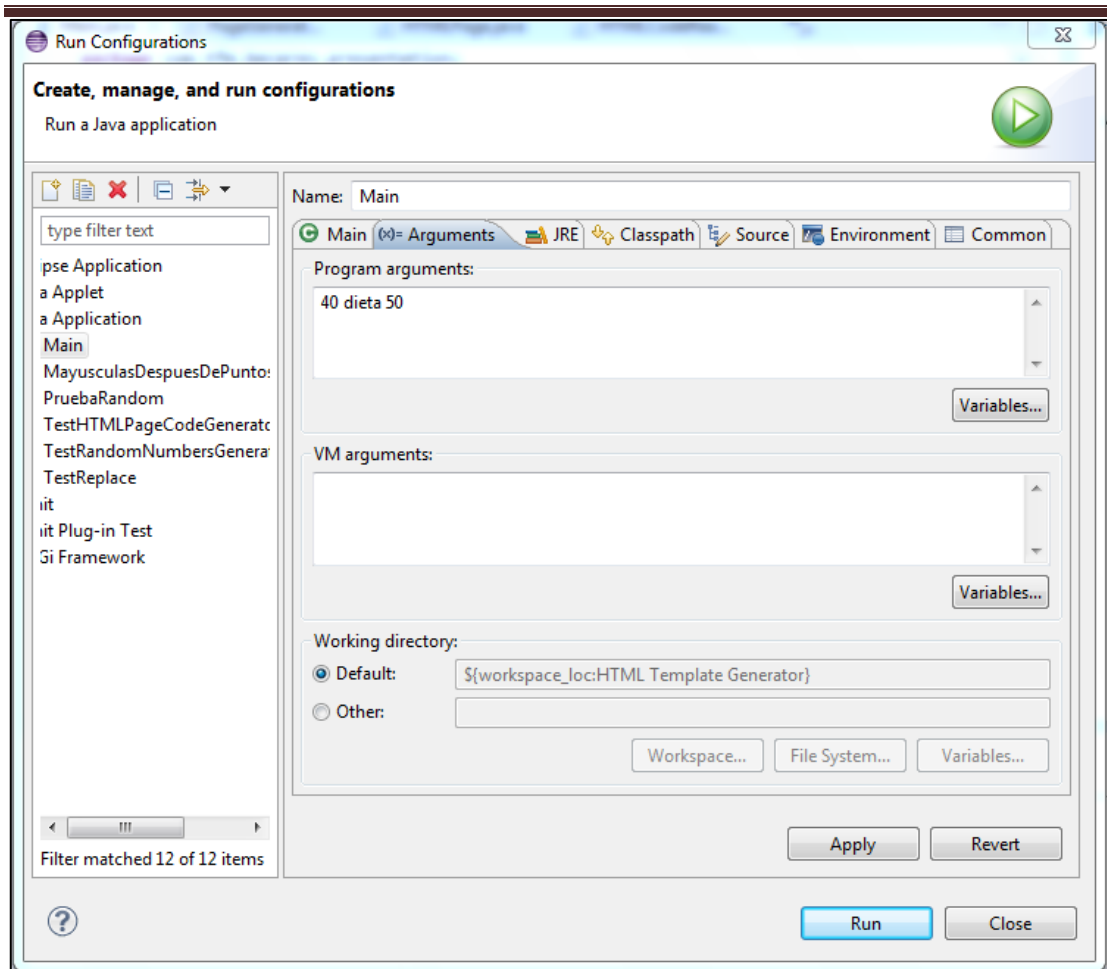


Ilustración 21 – Ejecución del prototipo desde Eclipse

Parámetros de ejecución del generador de páginas web:

- El primero es el número de páginas a generar
- El segundo la palabra clave
- El tercero el porcentaje de páginas web optimizadas que se generarán en la muestra.

Una vez ejecutado el prototipo tendremos resultados en el directorio donde se almacenan las páginas web generadas y en el directorio donde se almacenan los metadatos. Vamos a verlo por ejemplo para la palabra clave “Computador”:

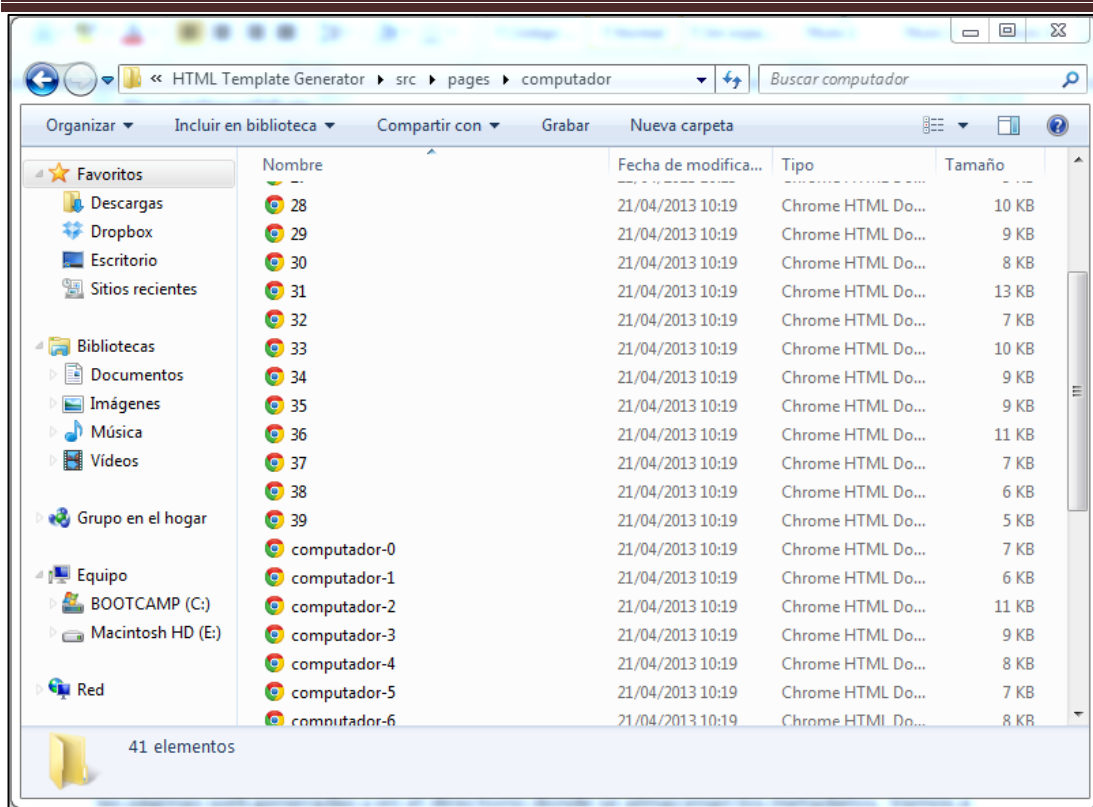


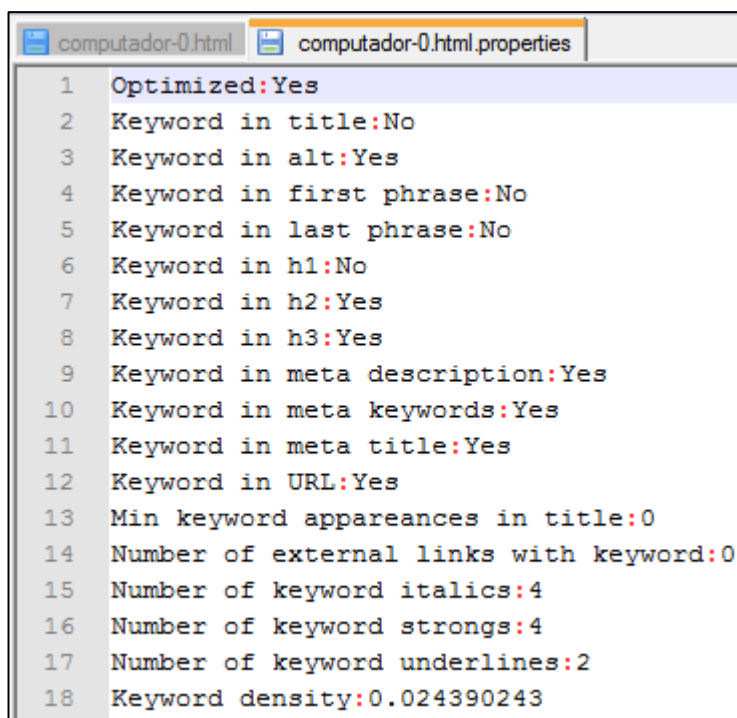
Ilustración 22 – Páginas generadas para la palabra clave “Computador”

Código generado para una página web de la palabra clave "Computador":

```
<!DOCTYPE HTML>
<html>
<head>
<title>
El mejor antivirus del mercado
</title>
<meta http-equiv="content-type" content="text/html; charset=UTF-8" />
<meta name="title" content="¿Qué partes forman una computadora?" />
<meta name="description" content="Computador - Significados en español y discusiones con el uso de 'computador'." />
<meta name="keywords" content="Memoria, procesador, RAM, computador" />
</head>
<body>
<h1>Mejora tu ordenador</h1>
<p>La unidad aritmético lógica o ALU es el dispositivo diseñado y construido para llevar a cabo las operaciones elementales como las operaciones aritméticas (suma, resta, ...), operaciones lógicas (Y, O, NO), y operaciones de comparación o relacionales.</p>
<h2>Aprende a programar tu computadora</h2>
<p>Las minicomputadoras son una clase de computadora multiusuario, que se encuentran en el rango intermedio del espectro computacional; es decir entre los grandes sistemas multiusuario (mainframes), y los más pequeños sistemas monousuarios (microcomputadoras, computadoras personales, o PC, etc.)</p>
<p>Los primeros <i>computador</i>es digitales, de gran tamaño y coste, se utilizaban principalmente para hacer cálculos científicos. ENIAC, uno de los primeros computadores, calculaba densidades de neutrón transversales para ver si explotaría la bomba de hidrógeno. El CSIR Mk I, el primer ordenador australiano, evaluó patrones de precipitaciones para un gran proyecto de generación hidroeléctrica. Los primeros visionarios vaticinaron que la programación permitiría jugar al ajedrez, ver películas y otros usos.</p>
<p>El disco duro es un sistema de grabación magnética digital, es donde en la mayoría de los casos reside el Sistema operativo de la ordenador. En los discos duros se almacenan los datos del usuario. En él encontramos dentro de la carcasa una serie de platos metálicos apilados girando a gran velocidad. Sobre estos platos se sitúan los cabezales encargados de leer o escribir los impulsos magnéticos.</p>
<p>Los primeros computadores digitales, de gran tamaño y coste, se utilizaban principalmente para hacer cálculos científicos. ENIAC, uno de los primeros computadores, calculaba densidades de neutrón transversales para ver si explotaría la bomba de hidrógeno. El CSIR Mk I, el primer ordenador australiano, evaluó patrones de precipitaciones para un gran proyecto de generación hidroeléctrica. Los primeros visionarios vaticinaron que la programación permitiría jugar al ajedrez, ver películas y otros usos.</p>
<p>La gente que trabajaba para los gobiernos y las grandes empresas también usó los <u>computador</u>es para automatizar muchas de las tareas de recolección y procesamiento de datos, que antes eran hechas por humanos; por ejemplo, mantener y actualizar la contabilidad y los inventarios. En el mundo académico, los científicos de todos los campos empezaron a utilizar los computadores para hacer sus propios análisis.</p>
<p>Otras impresoras, llamadas impresoras de red, tienen una interfaz de red interna (típicamente wireless o Ethernet), y que puede servir como un dispositivo para imprimir en papel algún documento para cualquier usuario de la red. Hoy en día se comercializan impresoras multifuncionales que aparte de sus funciones de impresora funcionan simultáneamente como fotocopidora y escáner, siendo éste tipo de impresoras las más recurrentes en el mercado.</p>
<p>La primera computadora portátil considerada como tal fue la Epson HX-20, desarrollada en 1981, a partir de la cual se observaron los grandes beneficios para el trabajo de científicos, militares, empresarios y otros profesionales que vieron la ventaja de poder llevar con ellos su computadora con toda la información que necesitaban de un lugar a otro.</p>
<p>Las minicomputadoras son una clase de computadora multiusuario, que se encuentran en el rango intermedio del espectro computacional; es decir entre los grandes sistemas multiusuario (mainframes), y los más pequeños sistemas monousuarios (microcomputadoras, computadoras personales, o PC, etc.)</p>
<p>La gente que trabajaba para los gobiernos y las grandes empresas también usó los <i>computador</i>es para automatizar muchas de las tareas de recolección y procesamiento de datos, que antes eran hechas por humanos; por ejemplo, mantener y actualizar la contabilidad y los inventarios. En el mundo académico, los científicos de todos los campos empezaron a utilizar los computadores para hacer sus propios análisis.</p>
<p>El sistema operativo es una especie de caja de herramientas lleno de utilerías que sirve para decidir, por ejemplo, qué programas se ejecutan, y cuándo, y qué fuentes (memoria o dispositivos E/S) se utilizan. El sistema operativo tiene otras funciones que ofrecer a otros programas, como los códigos que sirven a los programadores, escribir programas para una máquina sin necesidad de conocer los detalles internos de todos los dispositivos electrónicos conectados. En la actualidad se están empezando a
```

```
incluir en las distribuciones donde se incluye el sistema operativo, algunos programas muy usados, debido a que es ésta una manera económica de distribuirlos. No es extraño que un sistema operativo incluya navegadores de Internet, procesadores de texto, programas de correo electrónico, interfaces de red, reproductores de películas y otros programas que antes se tenían que conseguir e instalar separadamente.</p>
<p>Los altavoces se utilizan para escuchar los sonidos emitidos por el computador, tales como música, sonidos de errores, conferencias, etc. Altavoces de las placas base: Las placas base suelen llevar un dispositivo que emite pitidos para indicar posibles errores o procesos, o para indicar acciones para las personas con discapacidades visuales, como la activación del blog num, blog mayus.</p>
<p>Una Workstation es, en esencia, un equipo orientado a trabajos personales, con capacidad elevada de cálculo y rendimiento superior a los equipos PC convencionales, que aún tienen componentes de elevado coste, debido a su diseño orientado en cuanto a la elección y conjunción sinérgica de sus componentes.</p>
</body>
</html>
```

Metadatos generados para esa página web:



```
1 Optimized: Yes
2 Keyword in title: No
3 Keyword in alt: Yes
4 Keyword in first phrase: No
5 Keyword in last phrase: No
6 Keyword in h1: No
7 Keyword in h2: Yes
8 Keyword in h3: Yes
9 Keyword in meta description: Yes
10 Keyword in meta keywords: Yes
11 Keyword in meta title: Yes
12 Keyword in URL: Yes
13 Min keyword appareances in title: 0
14 Number of external links with keyword: 0
15 Number of keyword italics: 4
16 Number of keyword strongs: 4
17 Number of keyword underlines: 2
18 Keyword density: 0.024390243
```

Ilustración 23 – Metadatos generados para una página web de la palabra clave “Computador”

De esta forma generaremos todas las páginas y metadatos de las mismas para posteriormente subirlas a un servidor e indexarlas en los buscadores.

3.7 Indexación de las muestras creadas

Para llevar a cabo la indexación de páginas web es necesario que estas se encuentren en un lugar accesible para los buscadores. En este caso, el lugar escogido será un servidor web creado en el equipo asignado en el Máster de Ingeniería Web.

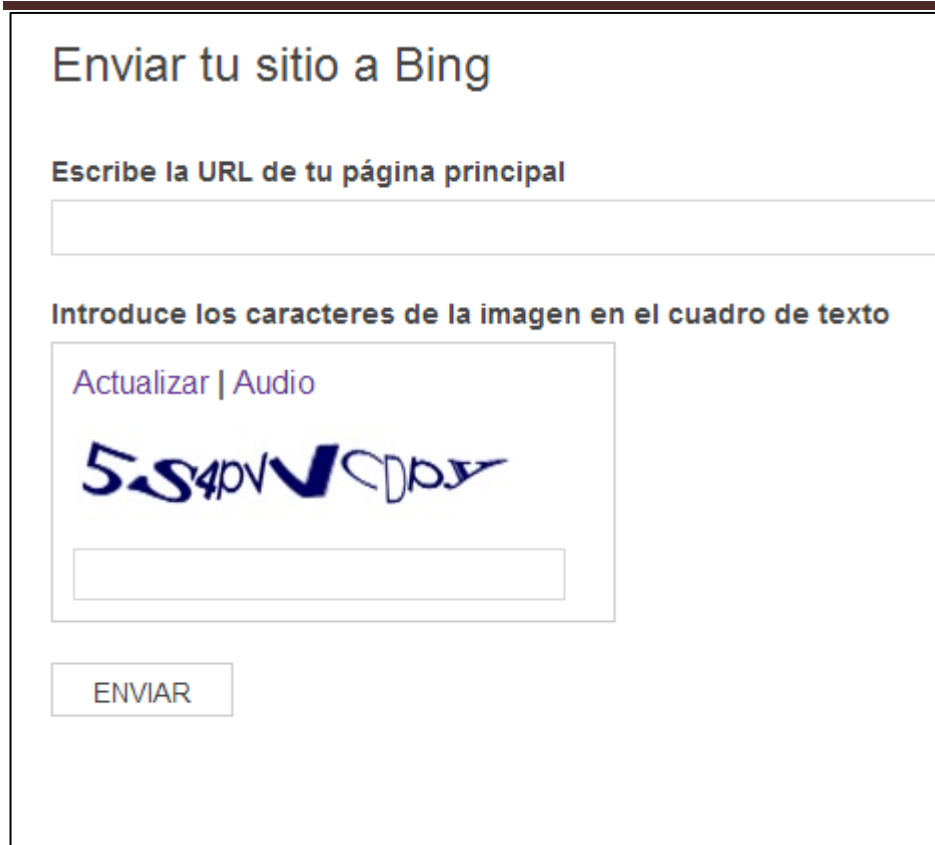
Las páginas se indexarán de la siguiente forma:

- Se creará un directorio para cada palabra clave, cuyo nombre será el de la propia palabra clave.
- Todas las páginas asociadas a una palabra clave, así como la página de enlaces a estas se insertarán en ese directorio.
- Los directorios colgarán de la raíz de la IP del equipo, quedando las direcciones con la forma '`http://[IP](:[Puerto])/[keyword]`'. Por tanto las páginas web tendrán como dirección siempre '`http://[IP](:[Puerto])/[keyword]/[nombre-de-archivo].html`'.
- Para que los motores de búsqueda indexen las páginas web se enviarán mediante el formulario creado para tal efecto en cada buscador las URLs de las páginas web de enlaces de cada palabra clave a estudiar, y además se crearán enlaces a estas desde páginas web ya existentes en los motores de búsqueda para que la araña de los buscadores las detecte. (Se crean enlaces a las páginas de enlaces, no a las páginas de contenido que se van a estudiar, así esto no afecta como factor externo). Además, en el caso de Google, se creará una cuenta de Google Webmaster Tools por si acaso, y en el caso de Bing de Herramientas para administradores web de Bing, las cuales ayudan a que estos motores de búsqueda indexen una URL.



The image shows a screenshot of the 'Herramientas para webmasters de Google' (Google Webmaster Tools) URL submission form. At the top, it says 'Herramientas para webmasters de Google' and has an 'Ayuda' (Help) dropdown. Below that, there is a paragraph of text: 'Cada vez que rastrea la Web, Google añade sitios nuevos al índice y actualiza los sitios existentes. Si tiene una nueva URL, comuníquenoslo a través de esta página. No añadimos todas las URL enviadas a nuestro índice y no podemos predecir la fecha en la que las URL aparecerán en el mismo ni garantizar que vayan a aparecer.' Below the text is a 'URL:' label followed by an empty text input field. Underneath the input field is a CAPTCHA image showing the words 'otter' and 'under' in a stylized font. Below the CAPTCHA is a label 'Escribe las dos palabras:' followed by a text input field. To the right of the input field is a 'reCAPTCHA' logo. At the bottom of the form is an 'Enviar solicitud' (Send request) button.

Ilustración 24 – Formulario de envío de páginas web a Google




Enviar tu sitio a Bing

Escribe la URL de tu página principal

Introduce los caracteres de la imagen en el cuadro de texto

Actualizar | Audio



ENVIAR

Ilustración 25 – Formulario de envío de URL a Bing

Cabe destacar que, cuando se envía una URL a los buscadores, estos no solo indexan si así lo creen oportuno la URL enviada, sino también todas las páginas web a las cuales enlaza esta URL, y es por eso que se han creado las páginas de índices para cada palabra clave, para no tener que enviar miles de URLs a los buscadores, algo que ralentizaría mucho el tiempo de desarrollo de esta investigación.

3.7.1 Publicación de páginas web en Internet

Las páginas web se alojarán en el PC proporcionado en el máster, en la IP 156.35.98.57. Así, accediendo desde cualquier navegador a la URL “http://156.35.98.57 podremos ver todo lo que se ha generado.

Una vez subido el sitio web, debemos proceder a su indexación por los buscadores. Llegados a este punto, nos hemos dado cuenta que Yahoo ya no utiliza su propio motor de búsqueda, sino que utiliza el de Bing, por lo que vamos a inscribir las páginas web en Google y Bing. Para ello lo añadimos a las herramientas para webmasters de cada uno de los buscadores y esperaremos 48 horas para ver los resultados obtenidos y que todo haya ido como se esperaba y las páginas web sean indexadas.

3.8 Cálculo de la métrica

El objetivo de esta investigación no es otro que obtener una métrica sobre los factores *on-page* en el posicionamiento de los buscadores, por lo que hace falta definir cómo va a calcularse esa métrica.

Una vez estén subidas e indexadas las páginas web, ya se puede empezar a trabajar en la obtención de resultados.

Para ello se realizarán una serie de búsquedas que contengan las palabras clave sobre las que se generaron páginas web en su forma exacta.

Para cada búsqueda se analizarán las páginas web que aparecen en los 5 primeros resultados, estudiando los siguientes factores:

- Que la palabra clave aparezca en el título.
- Número de apariciones de la palabra clave en el título.
- Etiquetas meta title, meta description y meta keywords.
- Si la palabra clave aparece en encabezados h1, h2 y h3.
- Densidad de la palabra clave en la página web.
- Si la palabra clave aparece en imágenes y sus textos alternativos.
- Si la palabra clave aparece en la primera y última frase de la página web.
- Número de negritas, cursivas y veces que aparece la palabra clave subrayada.
- Número de enlaces externos en los que aparece la palabra clave.
- Palabra clave en la URL de la página web.

Así, con todos estos factores, seremos capaces de elaborar una tabla para cada búsqueda en la que se relacione la posición en la que aparece la página web con un factor en concreto y a partir de ahí extraer conclusiones.

En función del número de veces que se repita el factor en las 5 primeras posiciones se le dará mayor o menor importancia a ese factor y, cuando se realicen todas las búsquedas sobre las diferentes palabras clave, se podrá hacer una media de la importancia de cada factor y, a partir de ahí, asignar un peso sobre el total a cada factor.

Así, la tabla con la que se analizará cada búsqueda tendrá la siguiente forma:

Factor / Posición	1	2	3	4	5
Palabra clave en el título					
Más de 1 aparición de la palabra clave en el título					
Palabra clave en meta-title					
Palabra clave en meta-description					
Palabra clave en meta-keywords					
Palabra clave en h1					
Palabra clave en h2					
Palabra clave en h3					
Palabra clave en primera frase					
Palabra clave en última frase					
Densidad de la palabra clave en %					
Palabra clave en negrita					
Palabra clave en cursiva					
Palabra clave subrayada					
Palabra clave en etiqueta alt					
Palabra clave en enlace externo					
Palabra clave en URL					

Tabla 1 – Plantilla para rellenar puntuaciones de los factores a estudiar para cada página web de una búsqueda

3.9 Obtención de resultados

Para obtener los resultados se ha realizado una búsqueda por cada palabra clave de la que se crearon muestras y vamos a ver los resultados y almacenarlos en una tabla. Como ya dijimos, van a analizarse los 5 primeros resultados de cada búsqueda en busca de patrones que sigan las mismas páginas, los cuales ya se anticipan a simple vista una vez se realiza la búsqueda.

Para realizar las diferentes búsquedas introducimos en el formulario de búsqueda de Google lo siguiente: “site:http://156.35.98.57 keyword”. De esta forma solo obtendremos resultados del conjunto de nuestras páginas web, nada más interfiere en los resultados, que era una de las premisas para llevar a cabo esta investigación.

Como se ha mencionado anteriormente, Yahoo utiliza el motor de búsqueda de Bing, por lo que ofrece los mismos resultados y no se estudiará. Pero además ha habido otro contratiempo, y es que Bing no ha indexado las muestras de páginas web creadas, por lo que solo se podrá analizar el comportamiento de Google que, por otra parte, es el principal motor de búsqueda.

3.9.1 Palabra clave “dieta”

Los cinco primeros resultados en Google se pueden ver en la siguiente imagen:



Ilustración 26 – 5 primeros resultados en Google para la palabra clave “dieta” de nuestro conjunto de muestras

Las páginas que aparecen en estos resultados son:

1. dieta-6.html -> Optimizada
2. dieta-27.html -> Optimizada
3. dieta-45.html -> Optimizada
4. dieta-20.html -> Optimizada
5. dieta-13.html -> Optimizada

Factor / Posición	1	2	3	4	5
Palabra clave en el título	1	1	1	0	1
Más de 1 aparición de la palabra clave en el título	0	1	1	0	0
Palabra clave en meta-title	1	0	1	0	1
Palabra clave en meta-description	0	0	1	1	1
Palabra clave en meta-keywords	1	1	1	1	1
Palabra clave en h1	1	1	1	0	1
Palabra clave en h2	0	1	1	1	0
Palabra clave en h3	0	1	1	0	1
Palabra clave en primera frase	1	1	1	1	0
Palabra clave en última frase	0	0	1	1	0
Densidad de la palabra clave en %	4	3.9	2.7	3.3	2.7
Palabra clave en negrita	0	1	1	1	0
Palabra clave en cursiva	1	0	1	1	2
Palabra clave subrayada	1	1	0	1	1
Palabra clave en etiqueta alt	1	1	0	0	0
Palabra clave en enlace externo	0	1	0	0	1
Palabra clave en URL	1	1	1	1	1

Tabla 2 – Tabla de resultados para la palabra clave “dieta”

3.9.2 Palabra clave “adelgazar”

Los cinco primeros resultados en Google se pueden ver en la siguiente imagen:

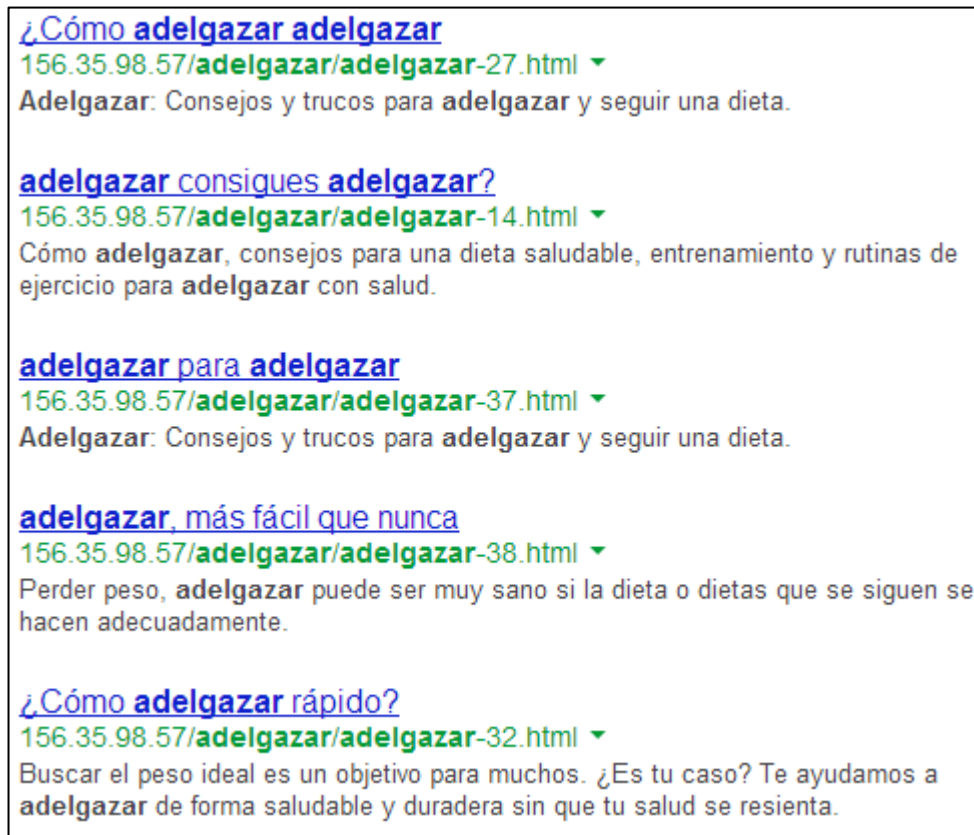


Ilustración 27 - 5 primeros resultados en Google para la palabra clave “adelgazar” de nuestro conjunto de muestras

Las páginas que aparecen en estos resultados son:

1. adelgazar-27.html -> Optimizada
2. adelgazar-14.html -> Optimizada
3. adelgazar-37.html -> Optimizada
4. adelgazar-38.html -> Optimizada
5. adelgazar-32.html -> Optimizada

Factor / Posición	1	2	3	4	5
Palabra clave en el título	1	1	1	1	1
Más de 1 aparición de la palabra clave en el título	1	1	1	0	0
Palabra clave en meta-title	1	1	0	1	1
Palabra clave en meta-description	1	1	1	1	1
Palabra clave en meta-keywords	1	1	1	1	0
Palabra clave en h1	1	0	1	1	1
Palabra clave en h2	1	0	0	1	0
Palabra clave en h3	0	1	0	0	0
Palabra clave en primera frase	1	1	1	0	0
Palabra clave en última frase	1	0	0	1	0
Densidad de la palabra clave en %	3.9	1.8	2.0	1.1	1.3
Palabra clave en negrita	0	1	1	1	1
Palabra clave en cursiva	0	0	0	1	1
Palabra clave subrayada	1	1	0	0	0
Palabra clave en etiqueta alt	1	0	0	0	0
Palabra clave en enlace externo	1	0	1	0	0
Palabra clave en URL	1	1	1	1	1

Tabla 3 - Tabla de resultados para la palabra clave "adelgazar"

3.9.3 Palabra clave “computador”

Los cinco primeros resultados en Google se pueden ver en la siguiente imagen:

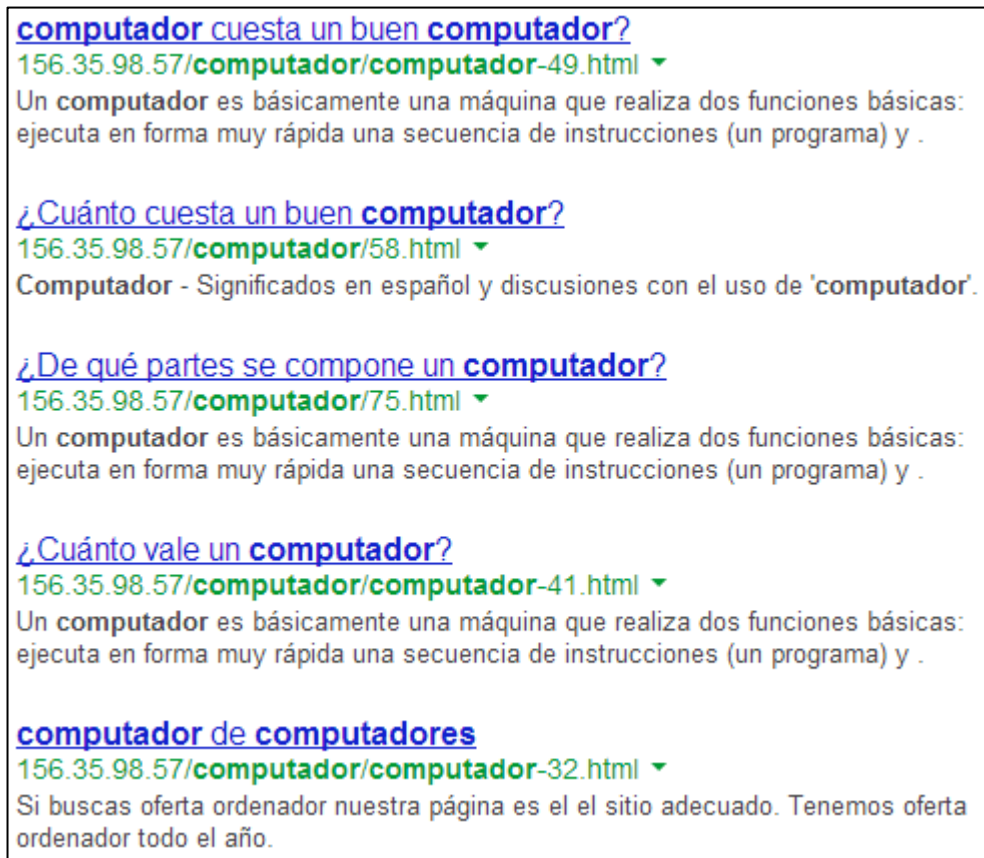


Ilustración 28 - 5 primeros resultados en Google para la palabra clave “computador” de nuestro conjunto de muestras

Las páginas que aparecen en estos resultados son:

1. computador-49.html -> Optimizada
2. 58.html -> No optimizada
3. 75.html -> No optimizada
4. computador-41.html -> Optimizada
5. computador-32.html -> Optimizada

Factor / Posición	1	2	3	4	5
Palabra clave en el título	1	1	1	1	1
Más de 1 aparición de la palabra clave en el título	1	0	0	0	1
Palabra clave en meta-title	0	0	0	1	1
Palabra clave en meta-description	1	1	1	1	0
Palabra clave en meta-keywords	1	0	0	1	0
Palabra clave en h1	0	1	1	1	1
Palabra clave en h2	1	0	0	0	1
Palabra clave en h3	0	0	0	1	0
Palabra clave en primera frase	0	1	0	1	1
Palabra clave en última frase	1	0	0	1	0
Densidad de la palabra clave en %	4.8	3.1	2.7	2	1.7
Palabra clave en negrita	0	1	0	0	1
Palabra clave en cursiva	1	0	0	1	1
Palabra clave subrayada	0	0	0	1	1
Palabra clave en etiqueta alt	1	0	1	0	0
Palabra clave en enlace externo	0	0	0	0	0
Palabra clave en URL	1	0	0	1	1

Tabla 4 - Tabla de resultados para la palabra clave "computador"

3.9.4 Palabra clave “car”

Los cinco primeros resultados en Google se pueden ver en la siguiente imagen:

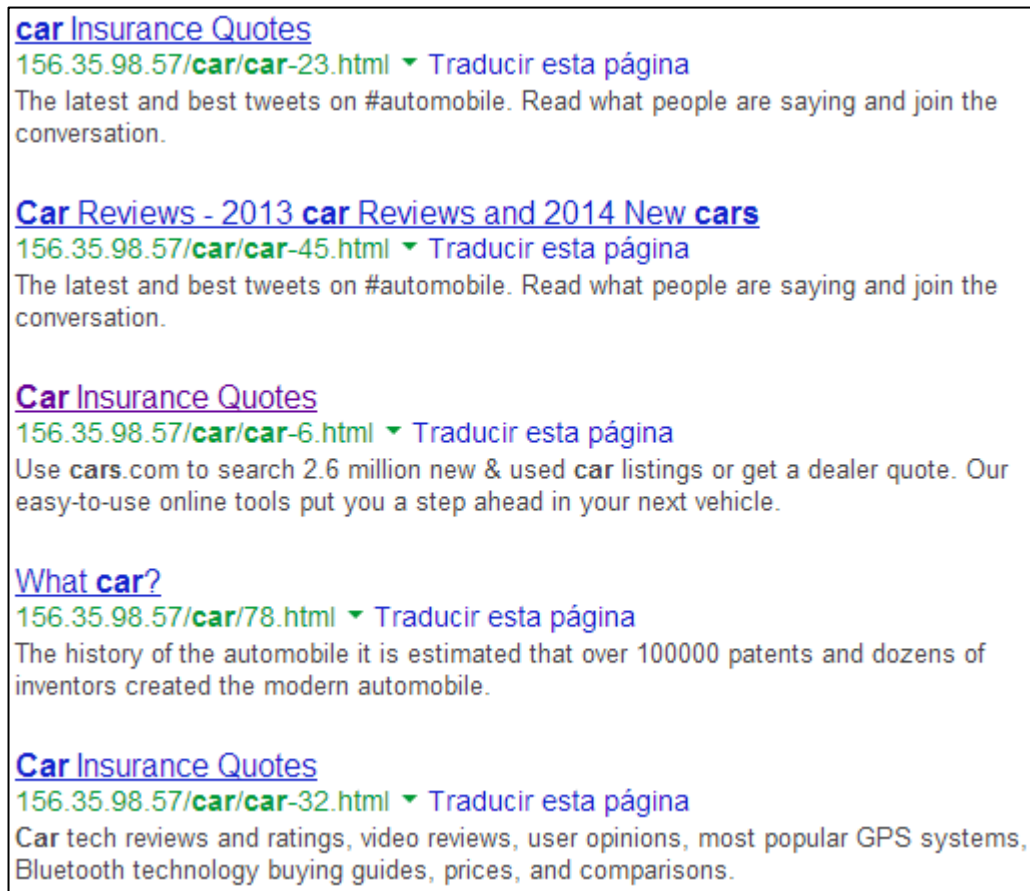


Ilustración 29 - 5 primeros resultados en Google para la palabra clave “car” de nuestro conjunto de muestras

Las páginas que aparecen en estos resultados son:

1. car-23.html -> Optimizada
2. car-45.html -> Optimizada
3. car-6.html -> Optimizada
4. 78.html -> No Optimizada
5. car-32.html -> Optimizada

Factor / Posición	1	2	3	4	5
Palabra clave en el título	1	1	1	1	1
Más de 1 aparición de la palabra clave en el título	0	1	0	0	0
Palabra clave en meta-title	0	1	0	0	0
Palabra clave en meta-description	0	0	0	0	0
Palabra clave en meta-keywords	0	0	0	1	1
Palabra clave en h1	1	1	1	1	1
Palabra clave en h2	1	1	0	0	0
Palabra clave en h3	1	0	0	0	1
Palabra clave en primera frase	1	1	1	0	0
Palabra clave en última frase	0	1	1	0	0
Densidad de la palabra clave en %	3.8	3.1	3.1	2.6	2.2
Palabra clave en negrita	0	0	0	1	1
Palabra clave en cursiva	0	1	0	0	0
Palabra clave subrayada	1	1	1	0	0
Palabra clave en etiqueta alt	0	0	0	0	0
Palabra clave en enlace externo	0	0	1	0	0
Palabra clave en URL	1	1	1	0	1

Tabla 5 - Tabla de resultados para la palabra clave "car"

3.10 Cálculo de la métrica

Para calcular la métrica se han utilizado los resultados obtenidos asignando el mismo peso a cada uno y sumando la densidad de la palabra clave al resultado final. De esta forma veremos si todos los factores tienen el mismo peso. Si fuera así, la suma total de la puntuación de cada factor en cada página daría como resultado un conjunto de totales ordenado de mayor a menor puntuación de izquierda a derecha.

Es decir, para que la métrica se haya obtenido, en todas las búsquedas realizadas cada resultado mejor posicionado que los demás debe tener mayor puntuación. El que se encuentra en la posición 1 debe tener mayor puntuación que el 2, el 2 que el 3 y así sucesivamente.

3.10.1 Primera iteración

Se puede ver que para la primera estimación asignando el mismo peso a cada factor no se cumple la condición de que todos los resultados estén ordenados por puntuación:

Dieta

Factor / Posición	1	2	3	4	5
Palabra clave en el título	1	1	1	0	1
Más de 1 aparición de la palabra clave en el título	0	1	1	0	0
Palabra clave en meta-title	1	0	1	0	1
Palabra clave en meta-description	0	0	1	1	1
Palabra clave en meta-keywords	1	1	1	1	1
Palabra clave en h1	1	1	1	0	1
Palabra clave en h2	0	1	1	1	0
Palabra clave en h3	0	1	1	0	1
Palabra clave en primera frase	1	1	1	1	0
Palabra clave en última frase	0	0	1	1	0
Densidad de la palabra clave en %	4	3,9	2,7	3,3	2,7
Palabra clave en negrita	0	1	1	1	0
Palabra clave en cursiva	1	0	1	1	1
Palabra clave subrayada	1	1	0	1	1
Palabra clave en etiqueta alt	1	1	0	0	0
Palabra clave en enlace externo	0	1	0	0	1
Palabra clave en URL	1	1	1	1	1
Totales	13	15,9	15,7	12,3	12,7

Adelgazar

Factor / Posición	1	2	3	4	5
Palabra clave en el título	1	1	1	1	1
Más de 1 aparición de la palabra clave en el título	1	1	1	0	0
Palabra clave en meta-title	1	1	0	1	1
Palabra clave en meta-description	1	1	1	1	1
Palabra clave en meta-keywords	1	1	1	1	0
Palabra clave en h1	1	0	1	1	1

Desarrollo de la investigación | Métrica de factores on-page en el posicionamiento de páginas web en los motores de búsqueda orgánicos.

Palabra clave en h2	1	0	0	1	0
Palabra clave en h3	0	1	0	0	0
Palabra clave en primera frase	1	1	1	0	0
Palabra clave en última frase	1	0	0	1	0
Densidad de la palabra clave en %	3,9	1,8	2	1,1	1,3
Palabra clave en negrita	0	1	1	1	1
Palabra clave en cursiva	0	0	0	1	1
Palabra clave subrayada	1	1	0	0	0
Palabra clave en etiqueta alt	1	0	0	0	0
Palabra clave en enlace externo	1	0	1	0	0
Palabra clave en URL	1	1	1	1	1
Totales	16,9	11,8	11	11,1	8,3

Computador

Factor / Posición	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Palabra clave en el título	1	1	1	1	1
Más de 1 aparición de la palabra clave en el título	1	0	0	0	1
Palabra clave en meta-title	0	0	0	1	1
Palabra clave en meta-description	1	1	1	1	0
Palabra clave en meta-keywords	1	0	0	1	0
Palabra clave en h1	0	1	1	1	1
Palabra clave en h2	1	0	0	0	1
Palabra clave en h3	0	0	0	1	0
Palabra clave en primera frase	0	1	0	1	1
Palabra clave en última frase	1	0	0	1	0
Densidad de la palabra clave en %	4,8	3,1	2,7	2,8	1,7
Palabra clave en negrita	0	1	0	0	1
Palabra clave en cursiva	1	0	0	1	1
Palabra clave subrayada	0	0	0	1	1
Palabra clave en etiqueta alt	1	0	1	0	0
Palabra clave en enlace externo	0	0	0	0	0
Palabra clave en URL	1	0	0	1	1
Totales	13,8	8,1	6,7	13,8	11,7

Car

Factor / Posición	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Palabra clave en el título	1	1	1	1	1
Más de 1 aparición de la palabra clave en el título	0	1	0	0	0
Palabra clave en meta-title	0	1	0	0	0
Palabra clave en meta-description	0	0	0	0	0
Palabra clave en meta-keywords	0	0	0	1	1
Palabra clave en h1	1	1	1	1	1
Palabra clave en h2	1	1	0	0	0
Palabra clave en h3	1	0	0	0	1
Palabra clave en primera frase	1	1	1	0	0
Palabra clave en última frase	0	1	1	0	0
Densidad de la palabra clave en %	3,7	3,1	3,1	2,6	2,2

Palabra clave en negrita	0	0	0	1	1
Palabra clave en cursiva	0	1	0	0	0
Palabra clave subrayada	1	1	1	0	0
Palabra clave en etiqueta alt	0	0	0	0	0
Palabra clave en enlace externo	0	0	1	0	0
Palabra clave en URL	1	1	1	0	1
Totales	10,7	13,1	10,1	6,6	8,2

Tabla 6 – Primera estimación de puntuación de las páginas web de cada búsqueda

A modo de resumen, los totales obtenidos son los siguientes:

Totales de “dieta”	13	15,9	15,7	12,3	12,7
Totales de “adelgazar”	16,9	11,8	11	11,1	8,3
Totales de “computador”	13,8	8,1	6,7	13,8	11,7
Totales de “car”	10,7	13,1	10,1	6,6	8,2

Tabla 7 – Resumen de la primera estimación de las puntuaciones

Como puede apreciarse, en ninguno de los casos aparecen los números ordenados de izquierda a derecha de mayor a menor, pero no están demasiado lejos. Sin embargo sí se pueden apreciar dos datos muy interesantes:

- Todas las páginas que aparecen en los 5 primeros resultados menos una tienen la palabra en el título, por lo que esto tiene que ser un factor de los más importantes.
- La densidad de la palabra clave es el factor que más se ajusta al orden de los resultados, teniendo casi siempre mayor densidad de la palabra clave los resultados mejor clasificados, y en el 100% de los casos la página en primera posición tiene mayor densidad que las demás. Por lo tanto, la densidad de la palabra clave es un factor muy importante.

3.10.2 Segunda iteración

Si nos fijamos en la palabra clave “adelgazar”, que es la única en la que los resultados aparecen bien ordenados salvo por una décima de diferencia entre el tercero y el cuarto, podemos sacar algunas conclusiones:

- Que la palabra clave aparezca en negrita o cursiva no parece tener mucho peso, puesto que el primer puesto no la tiene en negrita y el resto sí, los tres primeros no la tienen en cursiva y el resto sí. En cuanto a si la palabra aparece subrayada sí parece tener importancia, sin embargo en todas las directrices se da más importancia a las negritas que a los subrayados, así que puede ser simple casualidad, por lo tanto también vamos a considerar que esto tiene menos peso.
- Que la palabra clave aparezca más de una vez en el título puede ser importante ya que los tres primeros resultados cumplen ese requisito y, además, esto se puede extrapolar al resto de resultados obtenidos para el resto de búsquedas, ya que siempre al menos una página entre las 3 primeras posiciones cumple esto.

Si además buscamos también patrones en general, nos damos cuenta de lo siguiente:

- Que la palabra clave aparezca en la URL de la página es fundamental, ya que la mayoría de resultados cumplen este requisito.

Ahora, con todas estas conclusiones extraídas, vamos a elaborar nuevas tablas cambiando los pesos, y vemos los totales:

Totales de “dieta”	4,55	4,7	3,6	3,7	3,35
Totales de “adelgazar”	4,85	2,55	2,75	1,75	1,8
Totales de “computador”	5,55	3,5	3,1	3,5	2,35
Totales de “car”	4,2	3,8	3,6	2,9	2,65

Tabla 8 – Resumen de la segunda estimación de las puntuaciones

Ahora vemos que, con los cambios realizados, han cambiado algunas cosas. Para la palabra clave adelgazar ha empeorado la situación, pero para las otras tres ha mejorado notablemente.

3.10.3 Tercera iteración

Si seguimos ajustando índices fijándonos en los resultados que vamos obteniendo y, sobre todo, en la lógica que deben seguir los buscadores y aplicando los factores más importantes según todo el estado del arte leído, finalmente se ha conseguido asignar los pesos a cada factor de forma que cada página obtenga una puntuación más alta que todas las que están posicionadas peor en la búsqueda realizada.

Algunas de las directrices que se han aplicado son:

- Que la palabra clave aparezca en el título es fundamental para posicionar en las primeras posiciones una página web por esa palabra clave.
- La densidad de la palabra clave en la página es fundamental.
- La meta-descripción de la página también es muy importante ya que es el texto descriptivo que suelen mostrar los buscadores en las SERPs.

3.10.4 Tabla final de resultados

Dieta

Factor / Posición	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Palabra clave en el título	1,5	1,5	1,5	0	1,5
Más de 1 aparición de la palabra clave en el título	0	0,75	0,75	0	0
Palabra clave en meta-title	1	0	1	0	1
Palabra clave en meta-description	0	0	1	1	1
Palabra clave en meta-keywords	0,25	0,25	0,25	0,25	0,25
Palabra clave en h1	0,25	0,25	0,25	0	0,25
Palabra clave en h2	0	0,25	0,25	0,25	0
Palabra clave en h3	0	0	0	0	0
Palabra clave en primera frase	0,375	0,375	0,375	0,375	0
Palabra clave en última frase	0	0	0,375	0,375	0
Densidad de la palabra clave en %	14	13,65	9,45	11,55	9,45
Palabra clave en negrita	0	0	0	0	0
Palabra clave en cursiva	0	0	0	0	0
Palabra clave subrayada	0	0	0	0	0
Palabra clave en etiqueta alt	0,25	0,25	0	0	0
Palabra clave en enlace externo	0	0	0	0	0
Palabra clave en URL	0,5	0,5	0,5	0,5	0,5
Totales	18,125	17,775	15,7	14,3	13,95

Adelgazar

Factor / Posición	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Palabra clave en el título	1,5	1,5	1,5	1,5	1,5
Más de 1 aparición de la palabra clave en el título	0,75	0,75	0,75	0	0
Palabra clave en meta-title	1	1	0	1	1
Palabra clave en meta-description	1	1	1	1	1
Palabra clave en meta-keywords	0,25	0,25	0,25	0,25	0
Palabra clave en h1	0,25	0	0,25	0,25	0,25
Palabra clave en h2	0,25	0	0	0,25	0
Palabra clave en h3	0	0	0	0	0
Palabra clave en primera frase	0,375	0,375	0,375	0	0
Palabra clave en última frase	0,375	0	0	0,375	0
Densidad de la palabra clave en %	13,65	6,3	7	3,85	4,55
Palabra clave en negrita	0	0	0	0	0
Palabra clave en cursiva	0	0	0	0	0
Palabra clave subrayada	0	0	0	0	0
Palabra clave en etiqueta alt	0,25	0	0	0	0
Palabra clave en enlace externo	0	0	0	0	0
Palabra clave en URL	0,5	0,5	0,5	0,5	0,5
Totales	20,15	11,675	11,625	8,975	8,8

Computador

Factor / Posición	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Palabra clave en el título	1,5	1,5	1,5	1,5	1,5
Más de 1 aparición de la palabra clave en el título	0,75	0	0	0	0,75

Desarrollo de la investigación | Métrica de factores on-page en el posicionamiento de páginas web en los motores de búsqueda orgánicos.

Palabra clave en meta-title	0	0	0	1	1
Palabra clave en meta-description	1	1	1	1	0
Palabra clave en meta-keywords	0,25	0	0	0,25	0
Palabra clave en h1	0	0,25	0,25	0,25	0,25
Palabra clave en h2	0,25	0	0	0	0,25
Palabra clave en h3	0	0	0	0	0
Palabra clave en primera frase	0	0,375	0	0,375	0,375
Palabra clave en última frase	0,375	0	0	0,375	0
Densidad de la palabra clave en %	16,8	10,85	9,45	7	5,95
Palabra clave en negrita	0	0	0	0	0
Palabra clave en cursiva	0	0	0	0	0
Palabra clave subrayada	0	0	0	0	0
Palabra clave en etiqueta alt	0,25	0	0,25	0	0
Palabra clave en enlace externo	0	0	0	0	0
Palabra clave en URL	0,5	0	0	0,5	0,5
Totales	21,675	13,975	12,45	12,25	10,575

Car

<u>Factor / Posición</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Palabra clave en el título	1,5	1,5	1,5	1,5	1,5
Más de 1 aparición de la palabra clave en el título	0	0,75	0	0	0
Palabra clave en meta-title	0	1	0	0	0
Palabra clave en meta-description	0	0	0	0	0
Palabra clave en meta-keywords	0	0	0	0,25	0,25
Palabra clave en h1	0,25	0,25	0,25	0,25	0,25
Palabra clave en h2	0,25	0,25	0	0	0
Palabra clave en h3	0	0	0	0	0
Palabra clave en primera frase	0,375	0,375	0,375	0	0
Palabra clave en última frase	0	0,375	0,375	0	0
Densidad de la palabra clave en %	13,3	10,85	10,85	9,1	7,7
Palabra clave en negrita	0	0	0	0	0
Palabra clave en cursiva	0	0	0	0	0
Palabra clave subrayada	0	0	0	0	0
Palabra clave en etiqueta alt	0	0	0	0	0
Palabra clave en enlace externo	0	0	0	0	0
Palabra clave en URL	0,5	0,5	0,5	0	0,5
Totales	16,175	15,85	13,85	11,1	10,2

Tabla 9 – Tabla final de puntuaciones con los pesos finales asignados a cada factor

Métrica con la que se ha obtenido esta tabla a partir de la inicial:

Factor	Peso
Palabra clave en el título	1,5
Más de 1 aparición de la palabra clave en el título	0,75
Palabra clave en meta-title	1
Palabra clave en meta-description	1
Palabra clave en meta-keywords	0,25
Palabra clave en h1	0,25
Palabra clave en h2	0,25
Palabra clave en h3	0
Palabra clave en primera frase	0,375
Palabra clave en última frase	0,375
Densidad de la palabra clave en %	3,5
Palabra clave en negrita	0
Palabra clave en cursiva	0
Palabra clave subrayada	0
Palabra clave en etiqueta alt	0,25
Palabra clave en enlace externo	0
Palabra clave en URL	0,5

10

Tabla 10 – Pesos asignados a cada factor de la métrica

Como se puede apreciar, los factores más importantes por orden son:

1. Densidad de la palabra clave.
2. Que la palabra clave aparezca en el título.
3. Que la palabra clave aparezca en la etiqueta meta-title.
4. Que la palabra clave aparezca en la etiqueta meta-description.
5. Que la palabra clave aparezca más de una vez en el título.
6. Que la palabra clave aparezca en la URL de la página.

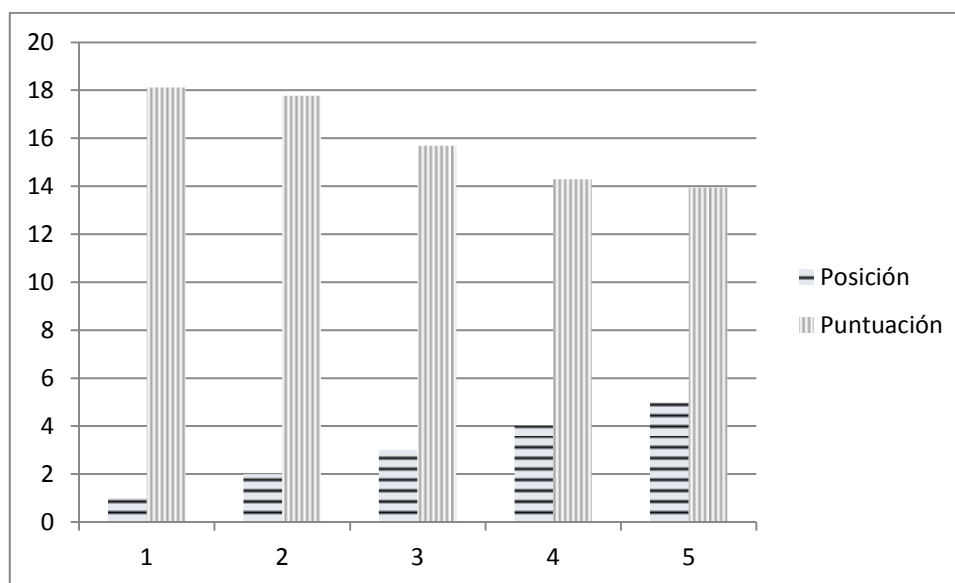
3.11 Gráficas de resultados

A continuación se presentan las gráficas obtenidas para cada búsqueda realizada con la puntuación obtenida por cada resultado y su posición.

En todas ellas, la columna denominada “Posición” indica el orden de aparición de la página web en el resultado de búsqueda. Así, la página web con posición igual a 1 es la que aparece primera en la búsqueda, la que tiene posición igual a 2 aparece segunda, y así sucesivamente. La columna denominada “Puntuación” indica el total de puntos obtenidos de la página en función de la métrica obtenida. Así, a mayor puntuación, menor posición debería tener una página web en los resultados de búsqueda y por tanto aparece antes que otras con menor puntuación.

3.11.1 Gráfica de puntuaciones obtenidas para la palabra clave “dieta”

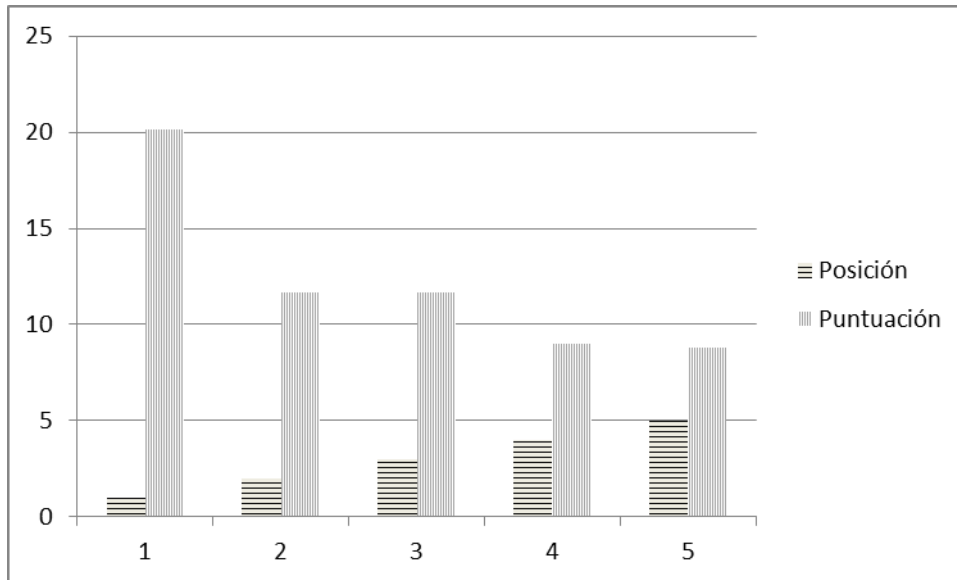
En esta gráfica podemos observar que apenas hay diferencias de puntuación entre posiciones, sobre todo entre la primera y segunda y entre la cuarta y quinta.



Gráfica 1 – Puntuaciones y posiciones para la palabra clave “dieta”

3.11.2 Gráfica de puntuaciones obtenidas para la palabra clave “adelgazar”

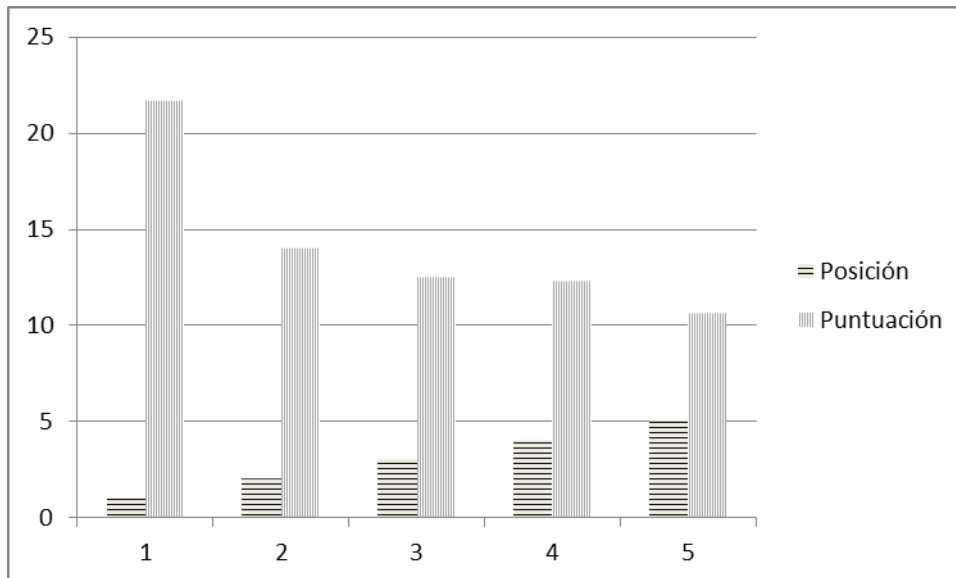
En esta segunda gráfica vemos como la página web de la primera posición ha obtenido una puntuación altísima en comparación con las demás, y que hay mucha igualdad entre la segunda y tercera posición, ya que entre ellas la diferencia de puntuación ha sido de 0,025 puntos.



Gráfica 2 - Puntuaciones y posiciones para la palabra clave “adelgazar”

3.11.3 Gráfica de puntuaciones obtenidas para la palabra clave “computador”

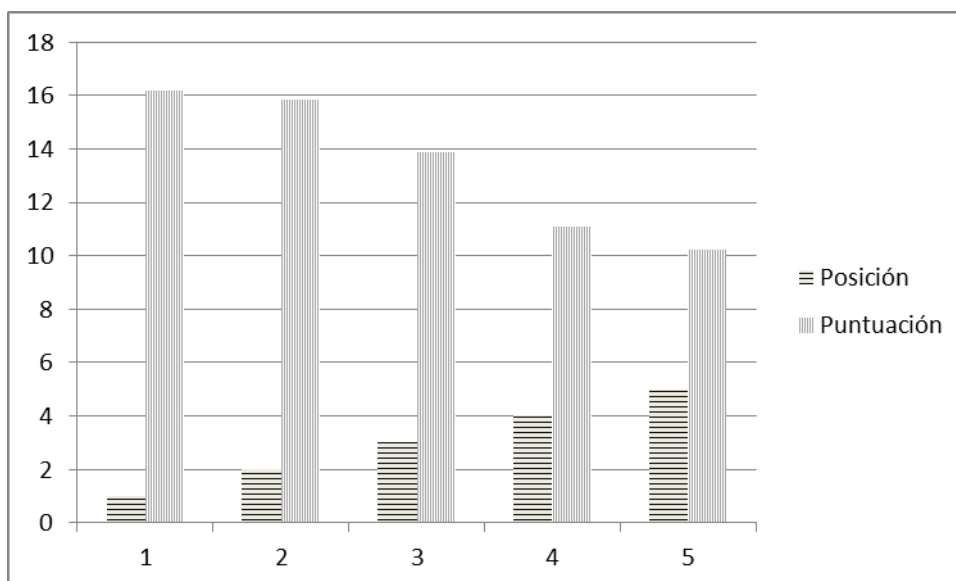
En la tercera gráfica podemos ver una tendencia similar a la de la segunda, con una primera posición muy destacada.



Gráfica 3 - Puntuaciones y posiciones para la palabra clave “computador”

3.11.4 Gráfica de puntuaciones obtenidas para la palabra clave “car”

En esta cuarta gráfica vemos una tendencia similar a la primera, con bastante igualdad entre puntuaciones de diferentes posiciones.



Gráfica 4 - Puntuaciones y posiciones para la palabra clave “car”

Parte IV - Conclusiones y trabajo futuro

Capítulo 4. Conclusiones y trabajo futuro

A continuación se exponen las principales conclusiones de la investigación realizada, así como las propuestas de trabajo futuro a seguir.

4.1 Hipótesis

La finalidad principal de esta investigación era demostrar que los motores de búsqueda utilizan algún tipo de métrica para ordenar los resultados de búsqueda, y determinar los factores que componen esta métrica, así como sus pesos.

Hipótesis: Los motores de búsqueda utilizan algún tipo de métrica para ordenar los resultados de búsqueda y es posible aproximarla gracias a la ingeniería inversa.

Tras el estudio realizado hemos podido concluir, tal y como se puede ver a continuación, que efectivamente la hipótesis se cumple y los motores de búsqueda utilizan algún tipo de métrica para ordenar los resultados de búsqueda y no lo hacen de forma aleatoria.

4.2 Objetivos

Todos los objetivos de la investigación se han cumplido de forma satisfactoria.

4.2.1 Objetivo principal

El objetivo primordial de la investigación era obtener una métrica con los factores que intervienen en la ordenación de resultados cuando se realiza una operación de búsqueda en un buscador. Además de los factores, obviamente, también era necesario determinar su peso.

Objetivo: Obtener una métrica con diferentes factores y pesos para cada factor a través de la cual podamos ordenar por puntuación los resultados de búsqueda.

El objetivo se ha cumplido al 100%. A lo largo de la investigación se han podido cumplir todos los pasos, desde la creación de muestras de páginas web hasta el estudio de los resultados y obtención de la métrica, pasando por la indexación de las páginas web estudiadas.

4.2.2 Generación de muestras de páginas web

Para la consecución del objetivo principal era necesario primero dar una serie de pasos que nos llevaran a construir el escenario adecuado para la investigación. Uno de estos pasos un objetivos era ser capaces de generar muestras de páginas web que cumplieran las condiciones establecidas en el apartado 3.2 del desarrollo de la investigación. La forma en que se han generado las páginas web está explicada en el apartado 3.5.

Objetivo: Generación de muestras de páginas web con las condiciones adecuadas para el estudio del posicionamiento de las mismas.

Gracias al prototipo desarrollado al inicio de la investigación se han podido crear muestras de páginas web del tamaño deseado y con la aleatoriedad necesaria para estudiar los diferentes factores que, juntos, conforman la métrica que se ha obtenido.

4.2.3 Indexación de páginas web

Este era un objetivo fundamental, ya que para poder estudiar el comportamiento de los buscadores tal y como queríamos era necesario que estos indexaran las páginas web generadas. Finalmente solo se han conseguido indexar en Google, tal y como se explica en el apartado 3.7.

Objetivo: Indexación de las diferentes muestras de páginas web generadas en los buscadores.

A pesar de que ha sido complicado debido a que Yahoo ya no existe como motor de búsqueda independiente y Bing no ha aceptado indexar las páginas web que se han enviado, al menos un buscador, Google, ha indexado las muestras que se han creado.

4.2.4 Obtención de la métrica que siguen los buscadores

Para terminar, gracias a la consecución de este objetivo hemos alcanzado el objetivo principal, que era demostrar que los buscadores efectivamente siguen una métrica con unos factores y pesos que llevan a la ordenación de los resultados en las búsquedas de forma no aleatoria. Pueden examinarse los resultados con más detalle en los apartados 3.9 y sucesivos.

Objetivo: Búsqueda de patrones y obtención, a partir de los mismos, de los pesos de cada factor propuesto en busca de la métrica.

El último y más complicado objetivo se ha cumplido, **se ha logrado obtener una métrica la cual nos ha permitido que, para todas las búsquedas realizadas, todos los resultados salen ordenados por puntuación de forma descendente en función de la posición de forma**

ascendente. Es decir, a menor posición (mejor posición en los resultados de búsqueda), mayor puntuación.

Los pesos obtenidos para cada factor han sido los siguientes:

Factor	Peso
Palabra clave en el título	1,5
Más de 1 aparición de la palabra clave en el título	0,75
Palabra clave en meta-title	1
Palabra clave en meta-description	1
Palabra clave en meta-keywords	0,25
Palabra clave en h1	0,25
Palabra clave en h2	0,25
Palabra clave en h3	0
Palabra clave en primera frase	0,375
Palabra clave en última frase	0,375
Densidad de la palabra clave en %	3,5
Palabra clave en negrita	0
Palabra clave en cursiva	0
Palabra clave subrayada	0
Palabra clave en etiqueta alt	0,25
Palabra clave en enlace externo	0
Palabra clave en URL	0,5

Tabla 11 – Pesos finales de cada factor de la métrica obtenida

4.3 Conclusiones acerca de los factores estudiados

En cuanto a los factores estudiados, se extraen las siguientes conclusiones:

1. Que la palabra clave aparezca en el título de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
2. El número de apariciones de la palabra clave en el título de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
3. Que la palabra clave aparezca en la meta descripción de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
4. Que la palabra clave aparezca la meta de título de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
5. Que la palabra clave aparezca en la meta de palabras clave de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
6. Que la palabra clave aparezca en el encabezado de nivel 1 de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
7. Que la palabra clave aparezca en el encabezado de nivel 2 de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
8. Que la palabra clave aparezca en el encabezado de nivel 3 de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **No se cumple.**
9. Que la palabra clave aparezca en la primera frase de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**

10. Que la palabra clave aparezca en la última frase de la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
11. El número de veces que la palabra clave aparece en negrita/cursiva/subrayada son factores que emplean los buscadores para obtener la relevancia de la misma. **No se cumple.**
12. El número de veces que la palabra clave aparece en enlaces externos es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
13. La densidad de la palabra clave en la página es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
14. Que la palabra clave aparezca en el texto alternativo de una imagen de la página web es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**
15. Que la palabra clave aparezca en la URL de la página web es un factor que emplean los buscadores para obtener la relevancia de la misma. **Se cumple.**

De los 15 factores estudiados, de tan solo 2 se ha deducido que no influyen en el posicionamiento de páginas web.

4.4 Recomendaciones para mejorar el posicionamiento de las páginas web

Por lo tanto, y siguiendo las conclusiones extraídas de la hipótesis y de la métrica obtenida, para posicionar una página web lo más alto posible en una búsqueda por una palabra clave, siguiendo las siguientes instrucciones habrá más probabilidades de conseguirlo:

- Es imprescindible que la palabra clave aparezca en el título de la página y si es más de una vez, mejor.
- La palabra clave debe aparecer en la etiqueta meta-title y la etiqueta meta-description de la página.
- La densidad de la palabra clave debe estar entre el 3% y el 5%, que son los valores que han obtenido las páginas mejor posicionadas en las muestras generadas.
- Si además de estas indicaciones la palabra clave aparece en el texto alternativo de una imagen, en las meta-keywords, en la URL de la página, en la primera y última frase de la web y en los encabezados h1 y h2 habrá más probabilidad de que la web aparezca más arriba en los resultados de búsqueda.

Además, con todos los cálculos realizados se pueden extraer varias conclusiones interesantes acerca de los datos obtenidos y del éxito de la investigación:

- El 90% de las páginas que aparecen en los cinco primeros resultados de búsqueda son optimizadas, lo cual quiere decir que, desde un principio, las hipótesis que se han seleccionado no estaban muy alejadas de la realidad en cuanto a cómo los motores de búsqueda ordenan los resultados en las SERPs.
- Algunos factores tienen más importancia que otros para los buscadores a la hora de ordenar los resultados en las páginas web, y los pesos que se han inducido con los datos de las hojas de cálculo lo demuestran.

- Hay factores que en principio considerábamos muy relevantes pero que al final, según la métrica obtenida, no lo han sido tanto, y viceversa.
- A juzgar por la diferencia que hay entre unas búsquedas y otras en cuanto a la puntuación obtenida de cada resultado comparándolo con el de su mismo índice en otras búsquedas cabe esperar que hay factores que no hemos tenido en cuenta y se nos escapan. Puesto que las páginas generadas no tienen sentido semánticamente hablando para un potencial lector de las mismas, seguramente cabe esperar que los buscadores hoy en día sean capaces de tener en cuenta la semántica de un documento.

No obstante hay que tener en cuenta que en esta investigación solo se han estudiado *factores on-page* de las páginas web, es decir, factores dentro de la propia web que influyen en su posicionamiento, por lo que el cumplimiento de estas recomendaciones no aseguraría estar mejor posicionado, ya que existen también los *factores off-page* como ya se ha mencionado anteriormente varias veces en el documento.

4.5 Difusión de los resultados

Como complemento a esta memoria de la investigación realizada se ha desarrollado un artículo, el cual se ha enviado a la revista “*Expert Systems with Applications*”. El scope de esta revista es bastante amplio, y aceptan trabajos realizados sobre el estudio, desarrollo o mejora de cualquier sistema, ya sea utilizado para fines científicos, gubernamentales o por la industria.

En el caso de esta investigación el sistema serían los motores de búsqueda orgánicos, empleados por grandes empresas de informática, como por ejemplo Google.

El índice de impacto de la revista actualmente es de 2.203 y en los últimos 5 años es de 2.455.

4.6 Trabajo futuro

Esta investigación no es más que un punto de partida sobre el que se pueden asentar futuros estudios acerca del posicionamiento en buscadores o SEO. Y es que dentro de esta investigación solo se han tenido en cuenta algunos de los factores *on-page* que en teoría utilizan los buscadores para asignar relevancia a las páginas web en los resultados de búsqueda.

El punto fuerte del estudio realizado es que han conseguido aislarse una serie de factores de todo lo demás, por lo que como trabajo futuro habría una serie de puntos que deberían afrontarse para ahondar más en cómo los buscadores ordenan las SERPs, como por ejemplo:

1. Aislar de alguna forma cada factor de los estudiados por separado y estudiar las diferencias entre páginas web idénticas salvo por ese factor.
2. Estudiar el tema de la semántica en los buscadores: comparar muestras que contenga resultados semánticamente correctos con otros que no lo son.

3. El SEO es algo que está en continua evolución, por lo que siempre hay que estar en busca de nuevos factores de los que se hable en artículos, foros, blogs y libros especializados para estudiarlos de forma similar a como se ha hecho con los seleccionados.
4. Incorporar al estudio realizado páginas web con enlaces entre sí, para estudiar así la influencia de los enlaces y los textos de los mismos en los resultados de búsqueda, uno de los factores que se consideran más importantes dentro del SEO.
5. Estudiar los *factores on-page* que no se han tenido en cuenta y que se cree que tienen influencia en la relevancia, como es la localización de la IP donde se aloja una página, el idioma, la antigüedad del dominio donde se aloja...
6. Estudiar cómo influye el tiempo de carga de las páginas web en los resultados de búsqueda.

Parte V - Referencias y Apéndices

Capítulo 5. Glosario

- **Backlink:** Enlace que apunta hacia un determinado sitio web.
- **Búsqueda orgánica:** Aquella en la que en sus resultados no aparecen resultados de pago.
- **CTR:** *Click-Through Rate*. El CTR es la frecuencia en la que gente clickea en resultados de búsqueda de tu sitio en relación con el número de impresiones de tus resultados de búsqueda.
- **Factores off-page:** Factores externos a una página web que influyen en su posicionamiento en los buscadores.
- **Factores on-page:** Factores internos de una página web que influyen en su posicionamiento en los buscadores.
- **SEM:** *Search Engine Marketing*. Conjunto de métodos, técnicas y herramientas empleadas para aparecer lo mejor posicionado posible en los resultados de pago de los buscadores con el menor coste por click y coste por impresión posible.
- **SEO:** *Search Engine Optimization*. Conjunto de métodos, técnicas y herramientas utilizadas para mejorar el posicionamiento de las páginas web en los motores de búsqueda orgánicos.
- **SERP:** *Search Engine Result Page*. Página de resultados de búsqueda de un buscador, donde se muestran las páginas web ordenadas por relevancia y, generalmente, con su título, descripción y URL.
- **Tasa de conversión:** relación entre el número de visitas que recibe una web y el número de operaciones comerciales o peticiones que se realizan en ella.

Capítulo 6. Referencias Bibliográficas

- Carlos, Juan. 2012. "200 Factores Del Algoritmo De Google." <http://www.forosdelweb.com/f64/analicemos-200-factores-parametros-del-algoritmo-google-1002851/>.
- Casares, Javier. 2008. *Guía De Referencia SEO*. <http://ojobuscador.com>.
- Casares, Javier. 2010. "Factores Internos (on Page) Del SEO." <http://javiercasares.com/seo/factores-internos-on-page/>.
- Codina, Lluís. 2007. "Posicionamiento Web : Conceptos y Ciclo De Vida." *Hypertext.net*.
- Codina, Lluís, and Mari Carmen Marcos. 2005. "Artículos Posicionamiento Web: Conceptos y Herramientas." *El Profesional De La Informacion* 14 (2) (March 1): 84–99. doi:10.3145/epi.2005.mar.01. <http://elprofesionaldelainformacion.metapress.com/openurl.asp?genre=article&id=doi:10.3145/epi.2005.mar.01>.
- Compete, Kantar Media. 2012. "Seeing Between the Lines ... of the Search and the Click."
- Cornello, Ignacio. 2010. "Factores on Page Del SEO." <http://expertoblog.com/factores-on-page-seo>.
- Cutts, Matt. 2011a. "Google Doesn't Use the Keywords Meta-tag in Web Search." <http://www.mattcutts.com/blog/keywords-meta-tag-in-web-search>.
- Cutts, Matt. 2011b. "What Is the Ideal Keyword Density of a Page?" http://www.youtube.com/watch?feature=player_embedded&v=Rk4qgQdp2UA.
- Google. 2012. *Guía Para Principiantes Sobre Optimización Para Motores De Búsqueda*.
- Internet World Stats. 2011. "Estadísticas De Internet En El Año 2011." <http://www.internetworldstats.com/stats.htm>. InternetWorldStats.com.
- Kobayashi, Mei, and Koichi Takeda. 2001. "Information Retrieval on the Web" 32 (2): 144–173.
- Larreina, Iñigo Arbildi. 2005. "Artículo Posicionamiento En Buscadores: Una Metodología Práctica De Optimización De Sitios Web." *El Profesional De La Informacion* 14 (2) (March 1): 108–124. doi:10.3145/epi.2005.mar.03. <http://elprofesionaldelainformacion.metapress.com/openurl.asp?genre=article&id=doi:10.3145/epi.2005.mar.03>.
- López Gómez, Miguel. 2009. *Libro SEO, Posicionamiento En Buscadores*. <http://libroseo.net>.
- MarketShare. 2012. "Datos Sobre Uso De Buscadores, Navegadores y Sistemas Operativos En Internet." <http://marketshare.hitslink.com/>.

- Martínez, Eduardo. 2012. "Contenido Duplicado: Una Pesadilla Para El SEO." <http://www.eduardomartinezblog.com/contenido-duplicado-malo-seo>.
- Peñarrolla, Montserrat. 2009. *Eyetracking: Search Marketing*.
- Search Engine Watch. 2013. "SEO & Keywords: Think Conversions, Not Rankings." <http://searchenginewatch.com/article/2258762/SEO-Keywords-Think-Conversions-Not-Rankings>.
- Su, Ao-Jan, Y. Charlie Hu, Aleksandar Kuzmanovic, and Cheng-Kok Koh. 2010. "How to Improve Your Google Ranking: Myths and Reality." *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (August): 50–57. doi:10.1109/WI-IAT.2010.195. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5616194>.
- Yalçın, Nursel, and Utku Köse. 2010. "What Is Search Engine Optimization: SEO?" *Procedia - Social and Behavioral Sciences* 9 (July 2009) (January): 487–493. doi:10.1016/j.sbspro.2010.12.185. <http://linkinghub.elsevier.com/retrieve/pii/S1877042810022901>.
- Zhang, Jin, and Alexandra Dimitroff. 2005. "The Impact of Webpage Content Characteristics on Webpage Visibility in Search Engine Results (Part I)." *Information Processing & Management* 41 (3) (May): 665–690. doi:10.1016/j.ipm.2003.12.001. <http://linkinghub.elsevier.com/retrieve/pii/S0306457303001122>.
- Zhao, Chengling, Jiaojiao Lu, and Fengfeng Duan. 2009. "Application and Research of SEO in the Development of Web2.0 Site." *2009 Second International Symposium on Knowledge Acquisition and Modeling*: 236–238. doi:10.1109/KAM.2009.69. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5362219>.
- Zhu, Cen, and Guixing Wu. 2011. "Research and Analysis of Search Engine Optimization Factors Based on Reverse Engineering." *2011 Third International Conference on Multimedia Information Networking and Security* (November): 225–228. doi:10.1109/MINES.2011.99. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6103760>.

Capítulo 7. Apéndices

En esta sección va a indicarse donde está cada elemento del trabajo en el CD-ROM entregable y se proporciona un índice alfabético del documento.

7.1 Contenido Entregado en el CD-ROM

En la raíz del CD-ROM se encuentra el archivo “Memoria.PDF”, correspondiente a este propio documento.

Además, el contenido entregado en el CD-ROM contiene dos carpetas:

1. Prototipo
2. Artículo

7.1.1 Prototipo

En esta carpeta se encuentran todos los ficheros necesarios para poder utilizar el prototipo de generación de páginas web. Dentro se encuentran tanto los archivos ejecutables como el código fuente.

Estructura de la carpeta:

- HTML Template Generator
 - bin
 - com
 - tfm
 - becares
 - bussiness
 - model
 - persistence
 - presentation
 - metadata
 - pages
 - texts
 - src
 - com
 - tfm
 - becares
 - bussiness
 - model
 - persistence
 - presentation
 - jsoup-1.7.1.jar: Librería necesaria para la ejecución del prototipo.

Las carpetas com.tfm.becares.* son los ejecutables y archivos de código fuente en las carpetas “bin” y “src” respectivamente.

La carpeta “texts” contiene los textos de muestra a partir de los cuáles se generarán páginas web.

La carpeta “metadata” es donde se almacenan los metadatos generados de cada página web.

La carpeta “pages” es donde se almacenan las páginas web generadas para cada palabra clave.

Dentro de las carpetas “texts”, “metadata” y “pages” hay subcarpetas donde se almacenan los ficheros para cada palabra clave. Ya vienen incorporadas las carpetas con los textos utilizados para los ejemplos del trabajo y con las páginas web generadas y metadatos utilizados en los cálculos de la investigación.

7.1.2 Artículo

En esta carpeta se incorporan todos los ficheros utilizados en la escritura del artículo enviado a la revista de investigación: el propio artículo, la covering letter, archivo de Highlights o puntos relevantes, tablas, gráficas y figuras. Además se incorpora el reporte generado automáticamente por el sistema de envío de artículos de la revista en PDF.

En el apartado 4.5 pueden verse los detalles del estado de envío/aceptación del artículo, la revista seleccionada, su scope y su índice de impacto.

7.2 Índice Alfabético

B

buscador, 3, 19, 20, 23, 39, 47, 50, 55, 56, 87, 114
buscadores, 3, 18, 21, 22, 23, 24, 25, 29, 30, 31, 33, 34, 37, 38, 39, 40, 41, 42, 43, 47, 48, 51, 52, 53, 55, 56, 57, 58, 60, 65, 72, 74, 86, 87, 88, 89, 103, 114, 115, 116, 117
Búsqueda, 5, 24, 114

C

Conclusiones, 111, 113, 115

F

Factores, 5, 21, 34, 35, 37, 48

H

Hipótesis, 22, 113

M

metadatos, 23, 25, 33, 58, 60, 61, 66, 69, 73, 83, 86, 126
métrica, 21, 22, 23, 24, 25, 39, 47, 48, 89, 100, 106, 107, 113, 114, 115, 116, 117

motores de búsqueda, 3, 17, 18, 21, 22, 24, 25, 29, 30, 33, 40, 41, 42, 47, 53, 55, 58, 87, 113, 116

O

on-page, 5, 21, 25, 34, 38, 40, 41, 42, 43, 48, 49, 53, 89, 117, 118

P

páginas web, 17, 18, 20, 21, 22, 23, 24, 25, 29, 30, 31, 33, 36, 37, 39, 40, 41, 42, 47, 48, 51, 52, 53, 54, 55, 56, 57, 58, 60, 61, 62, 63, 65, 66, 71, 72, 73, 74, 75, 82, 83, 87, 88, 89, 91, 102, 113, 114, 116, 117, 118, 125, 126
palabra clave, 22, 23, 25, 36, 40, 41, 42, 43, 48, 49, 50, 51, 52, 53, 54, 57, 58, 60, 61, 63, 65, 66, 71, 72, 73, 74, 75, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 115, 116, 126
posicionamiento, 21, 23, 24, 25, 30, 31, 33, 34, 36, 37, 39, 41, 42, 48, 52, 89, 114, 116, 117

S

Search Engine Optimization, 21
SEO, 3, 5, 7, 9, 20, 21, 29, 30, 31, 32, 33, 36, 38, 39, 40, 41, 52, 117, 118
SERP, 19, 20, 48, 50, 51, 53, 57, 58