



Universidad de Valladolid

Escuela de Ingeniería Informática

TRABAJO FIN DE GRADO

Grado en Ingeniería Informática
(Mención en Ingeniería de Software)

Optimización del posicionamiento de páginas web en buscadores

Autor:

D. Abel Velasco Pollino

Tutor:

D. Iván Santos Tejido

Resumen

Los motores de búsqueda se han convertido en el principal medio a través del cual millones de personas acceden a todo tipo de información. Con un público potencial tan elevado resulta de vital importancia lograr el mejor posicionamiento posible con el fin de lograr atraer al mayor número de visitantes. En este documento trataremos de comprender el funcionamiento de los motores de búsqueda, estudiaremos su evolución a lo largo del tiempo y analizaremos los factores que los motores de búsqueda emplean a la hora de establecer la posición de una página web en los resultados de búsqueda.

Posteriormente aplicaremos ese conocimiento sobre el sitio web del que es objeto este proyecto, realizando las modificaciones pertinentes que nos lleven a adaptar el sitio web a los requerimientos establecidos por los distintos factores. De esta forma lograremos optimizar la posición de las páginas del sitio web en los resultados de búsqueda.

Tabla de contenidos

1. Introducción	7
1.1. ¿Que es el SEO?	7
1.2. Motores de búsqueda	7
1.2.1. Funcionamiento básico	7
1.2.2. Páginas de resultados (SERPs)	10
1.2.3. Situación del mercado	15
1.3. Google: la referencia	18
1.3.1. Historial de actualizaciones	19
1.3.1.1. Actualizaciones menores	19
1.3.1.2. Actualizaciones mayores	22
1.4. Objetivos	29
1.5. Metodología	29
1.6. Estructura de la memoria	29
2. Factores SEO	31
2.1. Lista de Factores SEO	32
2.1.1. Factores SEO On-Page	33
2.1.1.1. Contenido	33
2.1.1.2. Arquitectura y aspectos técnicos	34
2.1.1.3. Código HTML	35
2.1.2. Factores SEO Off-Page	36
2.1.2.1. Confianza	36
2.1.2.2. Enlaces	37
2.1.2.3. Personalización	39
2.1.2.4. Redes sociales	39
2.1.3. Factores negativos	40
3. Proceso de optimización	43
3.1. Auditoria previa	43
3.1.1. Estructura y activos del sitio web	43
3.1.2. Estado de indexación	45
3.1.3. Análisis de competidores	46
3.1.4. Aspectos visuales que mejorar	47
3.1.5. Posicionamiento previo	48
3.2. SEO On-Page	49
3.2.1. Contenido	50
3.2.1.1. Contenido de calidad, fresco y único	50
3.2.1.2. Investigación de palabras clave	51

3.2.1.2.1. Aspectos previos.....	51
3.2.1.2.2. Técnicas	54
3.2.1.2.3. Herramientas	56
3.2.1.2.4. Proceso de investigación de palabras clave.....	60
3.2.1.3. Uso de las palabras clave.....	65
3.2.1.4. Contenido vertical.....	66
3.2.1.5. Gestión del contenido duplicado.....	66
3.2.2. Arquitectura y aspectos técnicos.....	67
3.2.2.1. Arquitectura óptima	68
3.2.2.1.1. Accesible para los motores de búsqueda.....	68
3.2.2.1.2. Estructura interna óptima	72
3.2.2.1.3. Proceso de optimización de la arquitectura	74
3.2.2.2. Formato de las URLs.....	80
3.2.2.3. Optimización para dispositivos móviles	83
3.2.2.4. Tiempo de carga.....	84
3.2.2.4.1. Habilitar la compresión.....	87
3.2.2.4.2. Especificar caché del navegador.....	88
3.2.2.4.3. Minificar CSS, HTML y JavaScript.	91
3.2.2.4.4. Optimización de las imágenes.....	92
3.2.2.4.5. Eliminar el JavaScript que bloquea la visualización.....	93
3.2.2.4.6. Optimizar la carga de CSS.....	94
3.2.2.5. HTTPS/SSL.....	95
3.2.2.6. Gestión del contenido duplicado	96
3.2.3. Código HTML.....	97
3.2.3.1. Etiquetas de título.....	98
3.2.3.2. Etiqueta de descripción	100
3.2.3.3. Etiquetas de encabezado	102
3.2.3.4. Nombre de archivo y atributo alt de las imágenes	104
3.2.3.5. Anchor text	104
3.2.3.6. Datos estructurados	106
3.2.3.6.1. Lenguajes de marcado.....	107
3.2.3.6.2. Tipos de datos estructurados	108
3.2.3.6.2.1. Información y estructura del sitio web	108
3.2.3.6.2.2. Contenido del sitio web.....	111
3.2.3.6.3. Herramientas y Testing	115
3.2.3.6.4. Resultados del marcado de datos estructurados.....	116
3.2.3.7. Etiquetas Open Graph y Twitter cards	117

3.2.3.8. Otros	119
3.2.3.8.1. Google Scholar.....	119
3.2.3.8.2. Corrección de los datos bibliográficos de los archivos PDF	123
3.3. SEO Off-page	125
3.3.1. Enlaces	125
3.3.2. Redes sociales	128
4. Seguimiento y análisis de los resultados.....	131
4.1. Herramientas	131
4.2. Metodología	132
4.3. Análisis de los resultados.....	134
4.3.1. Visión global	134
4.3.2. Estado de indexación	137
4.3.3. Casos concretos.....	141
5. Conclusiones.....	153
6. Bibliografía	157
7. Anexos	163

1. Introducción

1.1. ¿Que es el SEO?

La optimización en motores de búsqueda, más conocida por sus siglas en inglés SEO (Search Engine Optimization), es una disciplina englobada dentro del marketing online [\[1\]](#). Se centra en la mejora de la visibilidad de un sitio/página web en los resultados orgánicos (gratuitos) de los motores de búsqueda. El principal motivo es que un aumento de la visibilidad, en general, conlleva un aumento de las visitas, de los clientes potenciales y, por último, de los beneficios.

El SEO es un proceso que abarca los aspectos técnicos y creativos necesarios para mejorar el posicionamiento, impulsar el tráfico y aumentar la reputación de un sitio web. Existen multitud de elementos a considerar en un proyecto SEO, desde aspectos generales como el funcionamiento de los motores de búsqueda, hasta aspectos específicos como el número y calidad de los enlaces recibidos por una página web. A lo largo de este documento trataremos todos los aspectos de un proyecto SEO.

Cabe recordar que, aunque hablemos continuamente de optimización en motores de búsqueda, el SEO no se trata únicamente de hacer sitios web enfocados a los motores de búsqueda. Los sitios web también han de estar optimizados para las personas, pues son el consumidor final del contenido y una experiencia de usuario negativa puede echar por tierra todos los esfuerzos de un proyecto SEO.

Los continuos avances en la tecnología (smartphones, asistentes de voz...) han modificado la forma en la que los usuarios utilizan los motores de búsqueda. Estos avances, junto con el propio afán de los motores de búsqueda por mejorar tanto su habilidad para procesar el contenido de las páginas web, como la forma en la que se interpretan las consultas de los usuarios, hacen que los motores de búsqueda estén envueltos en un proceso de constante actualización de sus algoritmos. En el año 2013 Google llevó a cabo 890 cambios en el algoritmo [\[2\]](#). En el 2016, 1600 [\[3\]](#).

Cada vez que los motores de búsqueda llevan a cabo una gran actualización (algoritmo Panda, Penguin...) el panorama en el mundo del SEO cambia notablemente. Esto hace del SEO una disciplina en constante evolución, en la que un cambio en los algoritmos de los motores de búsqueda puede echar por tierra meses de trabajo.

1.2. Motores de búsqueda

Como hemos visto, un proyecto SEO centra su objetivo, en parte, sobre los motores de búsqueda. En este apartado explicaremos su funcionamiento básico y haremos un repaso de la situación actual del mercado.

1.2.1. Funcionamiento básico

El objetivo de los motores de búsqueda es proporcionar los mejores resultados posibles a sus usuarios, y esto pasará por ofrecer aquellos resultados que resulten más relevantes y actualizados. Para ello necesitan descubrir todas las páginas que se encuentran públicamente disponibles en la

Web, organizarlas en un índice y seleccionar aquellas que mejor se ajusten y sean más relevantes a la consulta realizada. A continuación, explicaremos el funcionamiento básico de un motor de búsqueda.

- Rastreo (Crawling)

El primer paso de este proceso es el rastreo/crawling de la Web, del que se encargan los robots/crawlers/rastreadores de los motores de búsqueda. Se trata de un software que se encarga de descubrir todas las páginas web que se encuentran disponibles públicamente. El robot de Google recibe el nombre de Googlebot, el de Bing recibe el nombre de Bingbot.

El proceso de rastreo comienza con una lista de URLs, obtenida de rastreos previos y de los archivos sitemap proporcionados por los webmasters. A medida que el robot va visitando los sitios web de la lista, detecta los enlaces que contiene y los emplea para descubrir otras páginas web. El software presta especial atención a las páginas nuevas, a los cambios en los sitios web existentes y a los enlaces muertos. De esta forma, y aprovechando la estructura de enlaces propia de la Web, los robots tienen acceso a billones de documentos que se encuentran interconectados.

Pero este no es el único cometido de los rastreadores. A medida que recorren cada una de las páginas recogen la información contenida en las misma y la almacenan en los servidores de los respectivos motores de búsqueda para su posterior indexación.

Los rastreadores emplean un proceso algorítmico para determinar qué sitios web hay que rastrear, la frecuencia de dicho rastreo y cuántas páginas se deben rastrear en cada sitio web. Cuando se trata de sitios web pequeños, con apenas unos miles de páginas, el proceso de rastreo se lleva a cabo de forma eficiente la mayor parte del tiempo. En cambio, cuando se trata de sitios web de gran tamaño es importante priorizar qué rastrear, cuándo y cuántos recursos del servidor que aloja el contenido pueden ser destinados al rastreo.

Los robots (Googlebot, Bingbot...) tienen como prioridad rastrear la Web, siempre y cuando esto no afecte a la experiencia de los usuarios que visiten el sitio web. Existe un “Límite de la tasa de rastreo” [\[4\]](#) y representa el número máximo de conexiones simultáneas que el robot puede emplear para rastrear el sitio web, así como el tiempo que debe esperar entre cada rastreo. La tasa de rastreo viene marcada por dos factores:

- *Respuesta ante el rastreo:* Si el sitio web responde a las peticiones con rapidez el límite sube, y se amplían el número de conexiones para el rastreo. Si el sitio web se ralentiza el límite baja y el robot rastrea menos páginas.
- *Límite establecido por el Webmaster:* Google Search Console permite a los Webmaster establecer el límite de rastreo. Destacar que un límite elevado no asegura un aumento del rastreo.

La actividad de rastreo también depende de la demanda, de forma que las URLs que son más populares tienden a ser rastreadas con mayor frecuencia para mantener el índice actualizado. Por otro lado, se intenta que las URLs obsoletas no permanezcan en el índice.

Los administradores de los sitios web tienen la capacidad de restringir el acceso de los robots al contenido de los mismos. Para ello emplearán el archivo robots.txt, para bloquear el acceso a directorios completos, o bien el atributo *rel="nofollow"* para bloquear el acceso a páginas y links concretos [5] [6].

- Indexación (Indexing)

Una vez que los rastreadores/crawlers han recorrido los sitios web, recogiendo el contenido de sus páginas, llega el momento de actualizar el índice. Para ello el sistema de indexación procesa el contenido de cada una de las páginas de forma similar a como lo hace el navegador web. Se toman nota de todos los aspectos clave contenidos en las páginas: las palabras clave y su localización en la página, la frescura del contenido, elementos contenidos en ellas (imágenes, videos...), enlaces salientes, información incluida en las etiquetas y atributos del código HTML...

El índice de Google contiene billones de páginas web y ocupa más de 100.000.000 de gigabytes de tamaño. Contiene una entrada para cada una de las palabras que aparecen en cada una de las páginas web indexadas. De esta forma cada vez que se indexa una página web, se añade esta página a cada una de las entradas correspondientes a las palabras que contiene.

Hasta 2010, año en el que se actualiza por completo el sistema de indexación de Google con la inclusión de un nuevo algoritmo con el nombre en clave de Caffeine [7], el índice estaba compuesto por varias capas, con diferentes tasas de refresco. La capa principal se actualizaba cada dos semanas aproximadamente. Para actualizar una capa entera era necesario analizar toda Web, lo que introducía un retraso significativo entre el momento en el que la página era encontrada y el momento en el que se incluía en el índice y estaba a disposición de los usuarios.

Con Caffeine se analiza la Web en pequeñas porciones y se actualiza en índice de forma continua. De esta forma, cuando se encuentran nuevas páginas o nuevo contenido en páginas ya indexadas, se pueden incluir directamente en el índice, reduciendo al mínimo el retraso que incluía el anterior sistema y proporcionando al usuario información fresca y actualizada [5] [6].

- Resultados

El último paso, una vez tenemos en índice, es proporcionar los resultados más útiles y relevantes ante las consultas de los usuarios. Para ello los sistemas de clasificación de los motores de búsqueda, compuestos por una serie de algoritmos, deben analizar los billones de páginas web contenidas en el índice y seleccionar sólo aquellas páginas que resulten relevantes y útiles para el usuario.

Este proceso comienza con una consulta, realizada por un usuario. En primer lugar, los motores de búsqueda analizan los términos empleados en las consultas para tratar de entender su significado, y de esta forma ofrecer mejores respuestas. Para ello emplean una serie de algoritmos que permiten mejorar la forma en la que se interpretan las consultas como, por ejemplo, el algoritmo de detección e interpretación de errores ortográficos, el sistema de sinónimos, ... Además, se trata de categorizar el tipo de información que se busca a través de las palabras empleadas. Por ejemplo, ¿Se trata de una consulta específica o general? El uso de determinadas palabras como "imágenes", "vídeos", "horarios" ..., puede indicar una búsqueda más específica.

Una vez que se ha interpretado correctamente la consulta del usuario se procede a explorar el índice en busca de aquellas páginas web cuyo contenido se ajuste mejor a los términos empleados en dicha consulta. Se analiza la frecuencia y la posición en la que aparecen dichos términos, ya sea en los títulos, encabezados (etiquetas <h1> - <h6>), enlaces, ... o en el propio cuerpo de la página. También se analiza la existencia de otros tipos de contenido (imágenes, videos, ...) que ofrezcan valor añadido al contenido principal y, de esta forma, hagan que la página web en cuestión resulte más relevante para el usuario. Por último, se comprueba que la página web está escrita en el mismo idioma que la consulta, con el fin de priorizar aquellas páginas que se ajusten al idioma empleado por el usuario.

Como resultado se obtiene una lista con aquellas páginas cuyo contenido es potencialmente relevante para los términos de la consulta. Para hacernos una idea, la lista de resultados de una consulta normal puede estar compuesta por cientos de miles e incluso millones páginas web que es necesario clasificar en función de su utilidad para los usuarios. Para ello los motores de búsqueda cuentan con algoritmos encargados de determinar dicha utilidad. Es aquí donde entran en juego los “famosos” factores de clasificación de los motores de búsqueda.

En el caso de Google, sus algoritmos analizan más de 200 factores para tratar de establecer qué página ofrece la mejor información. Estos factores van desde la frescura del contenido, el número de veces que aparecen los términos de búsqueda, el tiempo de carga, número y calidad de los enlaces ... hasta la experiencia de usuario. También se trata de evaluar la confianza y autoridad de las páginas en la materia relativa a la consulta, para ello se buscan aquellas páginas que reciben más visitas ante consultas similares, aquellas que reciben enlaces de sitios web relacionados con la materia, ... factores que indican el grado de confianza que los usuarios depositan en la información.

Por otro lado, existen numerosos sitios web que emplean técnicas de spam, como la repetición de palabras clave, la ocultación de contenido, compra de enlaces..., con el fin de mejorar su posición en los resultados de búsqueda, y que no resultan útiles para los usuarios, aportando una mala experiencia de usuario y un contenido pobre. Por ello los motores de búsqueda cuentan con algoritmos encargados de la detección y penalización de aquellos sitios web que empleen este tipo de técnicas.

Los motores de búsqueda también ofrecen resultados de búsqueda personalizados, para los que tienen en cuenta factores relativos al contexto en el que se realiza la búsqueda. Entran en juego la localización física del usuario, su historial de navegación y de búsqueda, y sus preferencias establecidas en el propio buscador.

Por último, los motores de búsqueda comprueban la diversidad de los resultados ofrecidos. En general, los motores de búsqueda tratan de ofrecer un conjunto de resultados diverso, en los formatos que mejor se adapten a la consulta realizada. El caso más habitual es de aquellas palabras clave que tienen más de un significado. En estos casos los motores de búsqueda modifican el orden “normal” de los resultados para incluir resultados que abarquen los diversos significados de dichas palabras clave [\[8\]](#) [\[9\]](#).

1.2.2. Páginas de resultados (SERPs)

Una vez que los motores de búsqueda han confeccionado la lista final de resultados solo falta un último paso, mostrar al usuario la página de resultados, también conocida como SERP (Search Engine

Result Page). En análisis de su formato y de la forma en la que los usuarios interactúan con ella es de vital importancia para entender los objetivos del SEO [10].

- Formato

Hasta el año 2007, todos los buscadores de carácter general eran en sí mismos motores de búsqueda verticales, centrados específicamente en la búsqueda de páginas web. Ese año Google introduce el concepto de Universal Search (Búsqueda Universal) [11], en el que se combinan resultados de diferentes formatos digitales como imágenes, video, noticias... junto con el resto de resultados de búsqueda. A lo largo del tiempo el resto de motores de búsqueda han incluido funcionalidades similares.

En la Figura 1 podemos ver un fragmento de la infografía “Mega-SERP” [12], elaborada por Moz, en la se pueden ver los distintos elementos que a lo largo del tiempo han formado parte de la página de resultados de Google.

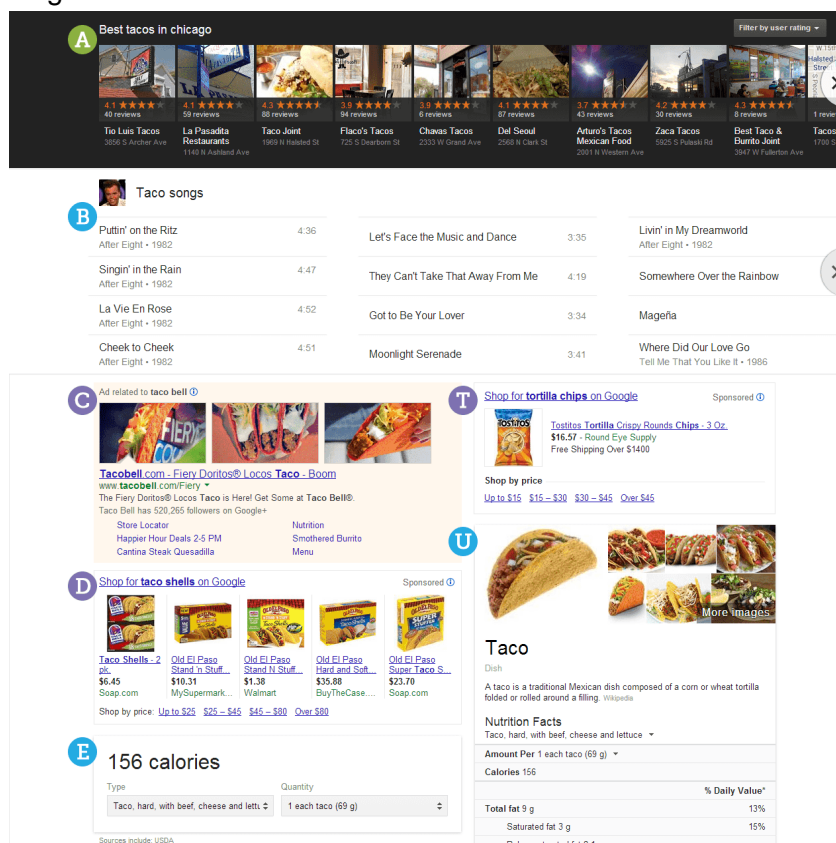


Figura 1. Infografía Mega-SERP. Moz.

Sin embargo, el formato “estándar” de una página de resultados es mucho más simple. La práctica totalidad de los motores de búsqueda presentan un formato similar. A continuación, veremos unos ejemplos de los formatos de las SERPs de Google y Bing, los elementos que tienen en común y aquellos en los que se diferencian:

En primer lugar, en la Figura 2, tenemos un ejemplo del formato que presenta la SERP de Google:

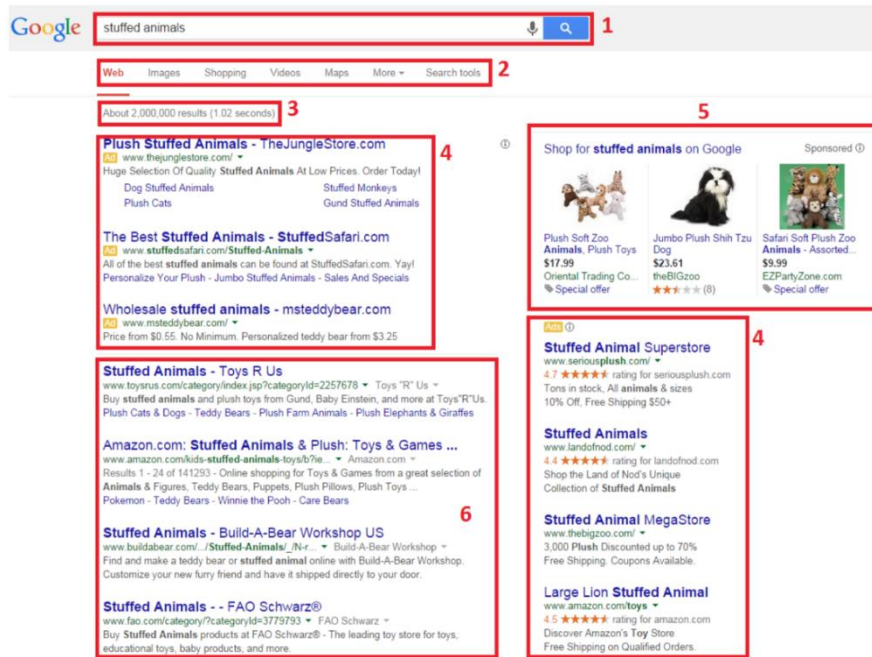


Figura 2. Formato de la SERP de Google.

Las distintas secciones marcadas en la imagen se corresponden con los siguientes elementos:

- Cuadro de búsqueda (1)
- Panel de navegación vertical y herramientas de búsqueda (2)
- Información sobre los resultados (número y tiempo empleado) (3)
- Anuncios de campañas de PPC (Pago por clic) (4)
- Elementos de la búsqueda universal (imágenes, videos, mapas, productos...) (5)
- Resultados naturales/orgánicos/algorítmicos (6)

Por otro lado, en la Figura 3 tenemos un ejemplo del formato de la SERP de Bing:

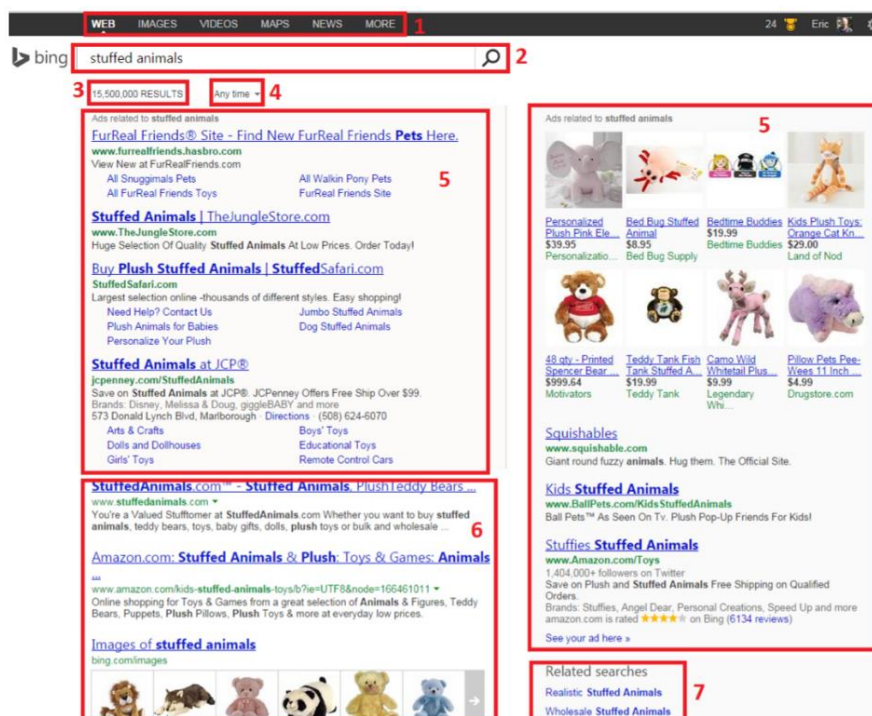


Figura 3. Formato de la SERP de Bing.

Las distintas secciones marcadas en la imagen se corresponden con los siguientes elementos:

- Panel de navegación vertical (1)
- Cuadro de búsqueda (2)
- Información sobre los resultados (3)
- Herramientas de búsqueda en función de la antigüedad del contenido (4)
- Anuncios de campañas de PPC (Pago por clic) (5)
- Resultados naturales/orgánicos/algorítmicos (6)
- Búsquedas relacionadas (7)

- Interacción de los usuarios

Una vez que hemos visto el formato de las páginas de resultados llega el momento de analizar la interacción de los usuarios con dichas páginas. Gracias a la inclusión de distintas evoluciones y cambios en su formato, como la Búsqueda Universal, la forma en la que los usuarios interaccionan con páginas de resultados ha cambiado.

A lo largo de los años se han llevado a cabo diferentes estudios con el propósito de analizar la forma en la que los usuarios interactúan con los motores de búsqueda. Para ello emplean técnicas de registro visual (eye-tracking) con el fin de determinar qué ven y en qué elementos se fijan los usuarios a la hora de realizar una búsqueda. Como resultado se obtienen mapas de calor que reflejan la atención que prestan los usuarios a las distintas secciones de una página de resultados.

En 2005, la empresa de investigación Enquiro (actualmente Mediative) publica su estudio “Google Golden Triangle Research: Eye Tracking Study” [13]. El mapa de calor incluido en el estudio refleja que los usuarios dedican su atención durante un mayor tiempo al contenido situado en la parte superior izquierda de la página de resultados. Conocida históricamente como el “Triángulo de oro”, tal y como se ve en la Figura 4.

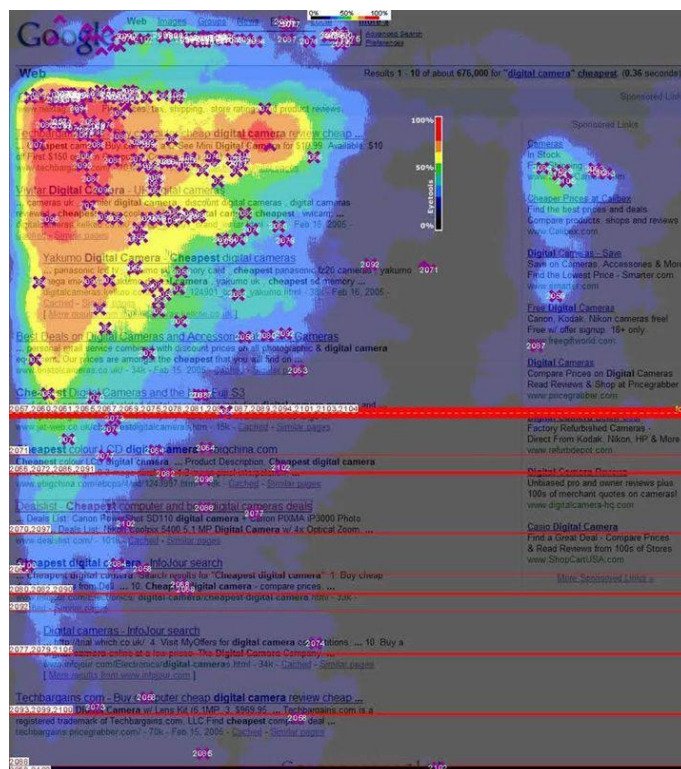


Figura 4. Resultados eye-tracking. Enquiro, 2005.

También queda patente la escasa atención que prestan los usuarios a los resultados de la parte baja en comparación a la que prestan a los resultados superiores. Además, el estudio refleja que los usuarios prestan más atención a las palabras en negrita, los títulos y las descripciones de los resultados orgánicos frente a los resultados de PPC (Pago por clic).

En 2007 Google introduce la Búsqueda Universal e introduce cambios en el formato de las páginas de resultados. Los distintos estudios llevados a cabo desde ese momento reflejan que, como resultado de estos cambios, se ha modificado la forma en la que los usuarios interactúan con los elementos de la página de resultados.

En 2014, Mediative publicó el estudio titulado “The Evolution of Google's Search Engine Results Pages and Their Effects on User Behaviour” [14]. El estudio refleja que el “Triángulo de oro” ha pasado a la historia. Los cambios en el formato de las páginas de resultados, con la inclusión de nuevos elementos (imágenes, vídeos, mapas, noticias, carrusel, knowledge graph...), han provocado que el primer resultado orgánico ya no se encuentre necesariamente en la parte superior izquierda. Esto, junto con la popularización de los smartphones, han hecho que los usuarios lleven a cabo una búsqueda mucho más vertical. Con ello, los elementos que ocupan posiciones más bajas reciben una mayor atención que la recibida anteriormente. En la Figura 5 tenemos dos mapas de calor que ejemplifican claramente el cambio en el comportamiento de los usuarios a la hora de interactuar con las páginas de resultados.



Figura 5. Resultados eye-tracking de Google SERP, 2014.

Existen otros métodos para medir la forma en la que los usuarios interactúan con las páginas de resultados. En este caso se trata de medir la Proporción de clics, conocido por sus siglas en inglés CTR (Click-through rate), que reciben los resultados dependiendo de su posición en la página de resultados. A lo largo de los años se han llevado a cabo diferentes estudios empleando dicho método. En la Tabla 1 figura un resumen de los datos recogidos por dichos estudios.

Tabla 1. Datos relativos a los distintos estudios de medición de CTR.

Posición	AOL '06 [15]	Enquiro '07 [16]	Chitika '10	Optify '10	Slingshot '11 [17]	Chitika '13 [18]	Catalyst '13 [19]	Caphyon '14 [20]
1	42,3	27,1	34,35	36,4	18,20	32,5	17,16	31,24
2	11,92	11,7	16,96	12,5	10,05	17,6	9,94	14,04
3	8,44	8,7	11,42	9,5	7,22	11,4	7,64	9,85
4	6,03	5,1	7,73	7,9	4,81	8,1	5,31	6,97
5	4,86	4,0	6,19	6,1	3,09	6,1	3,5	5,5
1 a 5	73,55	56,6	76,65	72,4	43,37	75,7	43,55	67,6
6	3,99	4,1	5,05	4,1	2,76	4,4	1,63	3,7 (6-10)
7	3,37	4,1	4,02	3,8	1,88	3,5	1,09	-
8	2,98	3,2	3,47	3,5	1,75	3,1	1,04	-
9	2,83	2,8	2,85	3,0	1,52	2,6	0,44	-
10	2,97	3,6	2,71	2,2	1,04	2,4	0,51	-
Top 10	89,69	73	95	89	52	92	48	71,3

Como se puede ver, los datos presentan variaciones sustanciales entre los distintos estudios. Esto se debe al uso de distintas metodologías por parte de cada uno de ellos. No es objeto de este proyecto entre en detalle de las metodologías aplicadas en cada uno de los estudios, estas pueden consultarse en [\[21\]](#).

Los datos de la Tabla 1 sí que nos permiten sacar una conclusión, pues son similares en todos los estudios realizados. Los 5 primeros resultados de las páginas de resultados reciben, de media, el 84% de los clics (orgánicos) que se producen en la primera página. También tenemos que, de media, el 76% de las búsquedas resultan con un clic en la primera página.

Los datos también ejemplifican la importancia de estar situado en la primera página de los resultados de búsqueda. De media, solo el 23,75% de las búsquedas pasan a la segunda página de resultados.

1.2.3. Situación del mercado

Actualmente, la oferta existente en el mercado de los motores de búsqueda es muy amplia. Dentro de este mercado existen dos tipos motores de búsqueda. Por un lado, tenemos los motores de

búsqueda de carácter general, como Google, Bing, Yahoo!, DuckDuckGo..., que ofrecen todo tipo de contenido. Por otro lado, tenemos los motores de búsqueda verticales, especializados en un sector en particular generalmente asociado al tipo de contenido que ofrecen. Entre ellos están Youtube, Yahoo! News, Bing Images, Expedia... Como podemos ver en los propios ejemplos, la mayoría de los motores de búsqueda cuenta con sus propios buscadores verticales, cuyo contenido integran en sus respectivas páginas de resultados.

Aunque la oferta es muy amplia los datos relativos a la cuota de mercado nos indican que hay un claro dominador, que concentra la mayoría del mercado de las búsquedas, tanto en escritorio como desde dispositivos móviles. Y ese no es otro que Google. En la Tabla 2 podemos ver los datos globales de los años 2015, 2016 y 2017, proporcionados de forma gratuita por la herramienta StatCounter [\[22\]](#), que toma datos de navegación a través de sus herramientas de análisis, instaladas en más de 2.5 millones de sitios web. Como se puede observar el dominio de Google es absoluto y va en aumento, tanto en búsquedas desde escritorio como en dispositivos móviles.

Tabla 2. Situación global del mercado de motores de búsqueda (Fuente: StatCounter)

	2017 (Hasta Abril)			2016			2015		
Motor	Global	Desktop	Móvil	Global	Desktop	Móvil	Global	Desktop	Móvil
Google	92,48%	88,74%	95,42%	92,01%	89,47%	95,03%	90,61%	88,66%	94,07%
Bing	2,82%	4,77%	1,23%	2,79%	4,36%	0,89%	3,02%	4,13%	1%
Yahoo!	2,13%	3,23%	1,2%	2,62%	3,14%	1,89%	3,4%	3,55%	2,82%
Baidu	1,05%	0,62%	1,4%	1,02%	0,64%	1,46%	0,96%	0,74%	1,37%
Yandex	0,75%	1,38%	0,30%	0,69%	1,06%	0,30%	0,71%	1,04%	0,26%
DDG!	0,15%	0,25%	0,06%	0,12%	0,15%	0,06%	0,11%	0,13%	0,05%
Otros	0,63%	1,01%	0,39%	0,75%	1,18%	0,37%	1,19%	1,75%	0,43%

Google domina el mercado de los motores de búsqueda, y en especial la búsqueda desde dispositivos móviles, que paulatinamente ha visto cómo han crecido sus cifras, y desde 2015 supera a la búsqueda desde escritorio.

Como se puede ver en las estadísticas de la Tabla 2, los tres grandes motores de búsqueda (Google, Bing, Yahoo!) acaparan la práctica totalidad del mercado. Esto ocurre así en la mayoría de países del planeta, pero existen excepciones. Países en los que los motores de búsqueda locales disputan e incluso arrebatan dicha hegemonía. En China, el principal motor de búsqueda es Baidu, superando ampliamente a los tres grandes motores de búsqueda. Algo más ajustada es la situación en Rusia, donde Yandex y Google se disputan el liderazgo. En Corea del Sur (Naver) y la República Checa (Seznam) sus respectivos motores de búsqueda locales ocupan la segunda posición tras Google, superando a Yahoo! y Bing.

Para completar la información repasaremos las cuotas de mercado de dos países que nos interesan especialmente en este documento. España (Tabla 3), por ser el país de origen del sitio web objeto este proyecto, y Estados Unidos (Tabla 4), marcado como el principal objetivo del contenido publicado en el sitio web.

Tabla 3. Situación del mercado de motores de búsqueda en España. (Fuente: StatCounter)

	2017 (Hasta Abril)			2016			2015		
Motor	Global	Desktop	Móvil	Global	Desktop	Móvil	Global	Desktop	Móvil
Google	95,9%	92,29%	99,14%	95,84%	93,74%	98,82%	95,74%	94,59%	98,09%
Bing	2,5%	5,15%	0,13%	2,47%	4,07%	0,19%	1,9%	2,64%	0,35%
Yahoo!	1,36%	2,14%	0,67%	1,44%	1,81%	0,9%	1,91%	2,18%	1,39%
Baidu	0,01%	0,01%	0,02%	0,02%	0,01%	0,02%	0,01%	0,01%	0,02%
Yandex	0,02%	0,03%	-	0,02%	0,02%	-	0,02%	0,02%	0,01%
DDG!	0,11%	0,20%	0,03%	0,1%	0,14%	0,03%	0,08%	0,11%	0,03%
Otros	0,10%	0,18%	0,01%	0,11%	0,21%	0,04%	0,34%	0,45%	0,11%

Tabla 4. Situación del mercado de motores de búsqueda en EE. UU. (Fuente: StatCounter)

	2017 (Hasta Abril)			2016			2015		
Motor	Global	Desktop	Móvil	Global	Desktop	Móvil	Global	Desktop	Móvil
Google	87,46%	80,7%	96,05%	85,65%	80,16%	94,35%	81,32%	76,95%	89,77%
Bing	6,33%	10,81%	0,62%	6,7%	10,2%	1,09%	8,49%	11,41%	2,86%
Yahoo!	5,28%	7,1%	2,93%	6,38%	7,68%	4,18%	8,71%	9,46%	6,87%
Baidu	0,07%	0,02%	0,06%	0,05%	0,05%	0,04%	0,05%	0,05%	0,07%
Yandex	0,02%	0,02%	-	0,01%	0,02%	-	-	-	-
DDG!	0,41%	0,5%	0,25%	0,34%	0,39%	0,22%	0,32%	0,37%	0,21%
Otros	0,43%	0,85%	0,09%	0,87%	0,50%	0,12%	1,11%	1,76%	0,22%

Como era de esperar el dominio de Google es aplastante, especialmente en España. En Estados Unidos, la situación está algo más equilibrada. De hecho, un estudio de cuota de mercado en dispositivos de escritorio y centrado en Estados Unidos, realizado por comScore en 2016, situaba la cuota de mercado de Bing en el 21%. En la Figura 6 podemos ver los datos completos.

comScore Explicit Core Search Share Report* (Desktop Only) February 2016 vs. January 2016 Total U.S. – Desktop Home & Work Locations Source: comScore qSearch			
Core Search Entity	Explicit Core Search Share (%)		
	Jan-16	Feb-16	Point Change
Total Explicit Core Search	100.0%	100.0%	N/A
Google Sites	63.8%	64.0%	0.2
Microsoft Sites	21.3%	21.4%	0.1
Yahoo Sites	12.4%	12.2%	-0.2
Ask Network	1.7%	1.6%	-0.1
AOL, Inc.	0.9%	0.9%	0.0

Figura 6. Situación del mercado de motores de búsqueda EE. UU., Febrero 2016. (Fuente: comScore)

Como hemos visto el dominio de Google sobre el mercado de los motores de búsqueda es prácticamente absoluto, y esto tiene consecuencias sobre el mundo del SEO. Una mayor cuota de mercado implica un mayor número de usuarios y, por tanto, un mayor número de clientes potenciales. Por ello, Google centra la atención de todo proyecto SEO. El seguimiento y la documentación de sus cambios es infinitamente más amplia que la de cualquiera de sus rivales.

Por último, destacaremos que no todos los motores obtienen sus resultados realizando su propio rastreo de la Web. El caso más relevante es el de Yahoo!. A partir del año 2010 los datos que muestra en sus páginas de resultados proceden de Bing. En Octubre de 2015 Yahoo! y Google cerraron un acuerdo en virtud del cual Google proporciona servicios a Yahoo! entre los que se incluyen anuncios, resultados de búsqueda de carácter general y de imágenes [23]. El acuerdo no se extiende a Europa debido a sus leyes antimonopolio.

En el caso de DuckDuckGo, el motor de búsqueda que apuesta por la privacidad de sus usuarios y no rastrea su comportamiento, obtiene sus resultados de alrededor de 400 fuentes de datos entre las que se encuentran otros motores de búsqueda, como Yahoo!, Yandex y Bing, con los que tiene acuerdos de colaboración [24].

En este proyecto se tomará como referencia a Google debido a su posición dominante en el mercado y a su abundante documentación. Además de Google, se llevará a cabo un seguimiento de los resultados de Bing y de Yandex, con el fin de comparar la incidencia de los cambios realizados.

1.3. Google: la referencia

Como hemos visto en la Sección 1.2, dentro del mundo del SEO Google es la referencia. En gran parte esto se debe a su posición en el mercado, que centra la atención de toda la industria y por ello es el motor de búsqueda cuyo comportamiento está más estudiado y documentado. Pero Google también es la referencia a nivel de algoritmo, tanto para el resto de motores de búsqueda, que terminan adoptando algunos de los cambios incluidos por Google, como para los webmasters de todo el mundo. Cuando cambia el algoritmo suele provocar cambios en los rankings y en la forma en la que se aplica el SEO.

En 2014 Google admitió que durante el año anterior llevó a cabo un total de 890 actualizaciones en su motor de búsquedas [2]. En 2016 el número creció hasta las 1600 mejoras [3]. No todas ellas implican cambios en profundos en los algoritmos, algunas de ellas hacen referencia a cambios en la interfaz, introducción de nuevas funcionalidades (Knowledge Graph, función autocompletar) ... De hecho, sólo tenemos noticias de un número bastante reducido, aquellas que supusieron un cambio perceptible. De otras muchas solo existen rumores y elucubraciones.

1.3.1. Historial de actualizaciones

A continuación, veremos una relación de las actualizaciones del algoritmo que han hecho evolucionar al SEO [25] [26]. Dentro de la importancia de dichas actualizaciones, las dividiremos en dos grupos, en función de su tipología, su incidencia inmediata y de las penalizaciones que acarrearón... Las actualizaciones “menores” están relacionados con cambios estéticos, nuevas funcionalidades... Las actualizaciones “mayores” están asociados a grandes cambios en los algoritmos.

1.3.1.1. Actualizaciones menores

Se trata de cambios importantes, pues fueron modelando la industria del SEO, pero cuya incidencia en los rankings no fue inmediata, sino más bien a largo a medio-largo plazo. Algunos de ellos son cambios estéticos, otros añaden funcionalidades, otros añaden nuevos factores a tener en el SEO...

A continuación, haremos una pequeña descripción de cada una de estas actualizaciones del algoritmo. Nos referiremos a ellas por los nombres en clave que les fueron asignados, tanto por Google como por la industria SEO.

- PageRank

Se trata de un algoritmo que aparece en los mismos orígenes de Google, en la tesis que Sergey Brin y Larry Page (Fundadores de Google) escribieron cuando estudiaban en la Universidad de Stanford [27]. PageRank es un algoritmo encargado de medir la importancia y la relevancia de un documento, para ello establecer un coeficiente (denominado PageRank) a cada página web en función de los enlaces que recibe y de la calidad de los mismos. A continuación, describiremos su funcionamiento básico.

En primer lugar, destacar que todas las páginas cuentan con una pequeña cantidad innata de PageRank. Las páginas aumentan su PageRank a medida que reciben enlaces de otras páginas web. Por tanto, cada página web tiene capacidad de pasar PageRank a otras páginas web. La cantidad de PageRank que una página puede pasar es menor que el PageRank con el que cuenta. Cuando una página únicamente enlaza a otra página web esta última recibe todo el PageRank del que dispone para pasar.

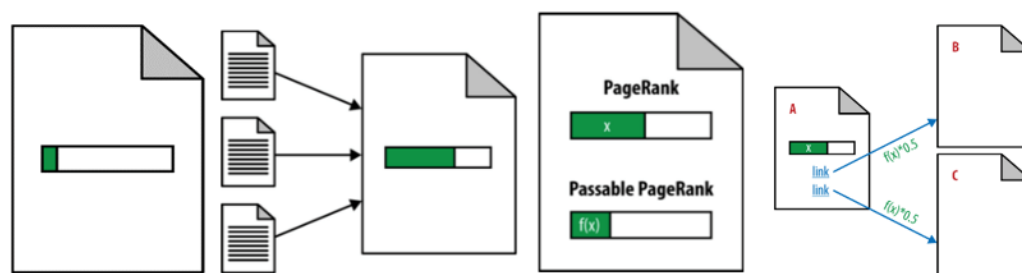


Figura 7. Funcionamiento básico de PageRank.

Sin embargo, cuando el número de páginas es mayor, el PageRank disponible para pasar se dividirá entre el número de enlaces salientes. Por ese motivo resulta conveniente que una página web no cuente con un excesivo número de enlaces salientes, pues la relevancia que transmite a las páginas enlazadas se verá diluida.

Este es un ejemplo muy simple de su funcionamiento, tal y como refleja la Figura 7, a medida que el número de enlaces aumenta y estos se empiezan a cruzar entre las distintas páginas, estableciendo dependencias, aumenta considerablemente la dificultad de su cálculo.

A lo largo de los años, y tras múltiples modificaciones, Google ha cambiado la forma en la que considera los enlaces y, por tanto, el algoritmo actual no está basado en el PageRank tal y como fue definido en su origen.

- **Florida (Noviembre 2003)**

Se trata de la primera actualización que puso al mundo del SEO en el mapa. Atacaba directamente a las más que cuestionables técnicas SEO de finales de los 90s, como el keyword stuffing, e hizo que muchos sitios web perdieran su posicionamiento.

- **Austin (Enero 2004)**

Continúa con el trabajo que empezó Florida. Elimina técnicas engañosas del SEO On-page, como el contenido oculto y el meta-tag stuffing.

- **Nofollow (Enero 2005)**

Con el fin de combatir el spam Google, Yahoo y Microsoft unen sus fuerzas para incluir el atributo "nofollow". Su utilización en un enlace nos permite indicar al motor de búsqueda que, en el momento de su rastreo, no siga dicho enlace. De esta forma se puede proteger a un sitio web de técnicas de spam que consisten en saturar las secciones de comentarios con enlaces a los sitios web que se desean posicionar.

- **Personalized Search (Junio 2005)**

Google comienza a hacer uso de los historiales de navegación de los usuarios para ofrecer resultados personalizados.

- **XML Sitemaps (Junio 2005)**

Permite a los administradores web enviar los archivos sitemap, en formato XML, directamente a Google a través de las Herramientas de Administrador. De esta forma los administradores tienen una pequeña influencia sobre el proceso de rastreo y de indexación.

- **Universal search (Mayo 2007)**

Google comienza a integrar los resultados de los motores de búsqueda vertical, como News, Video, Images, ... en los resultados de búsqueda tradicionales, cambiando para siempre el aspecto de las SERPs.

- **Google Suggest (Agosto 2008)**

Google cambia su página de inicio e incluye Suggest. El cuadro de búsqueda se transforma en una lista desplegable en la que aparecen sugerencias a medida que el usuario incluye su consulta. Más tarde Suggest derivaría en Google Instant.

- **Social signals (Diciembre 2010)**

Google y Bing confirman la utilización de “señales sociales” para determinar el ranking, incluyendo datos de Twitter y Facebook [\[28\]](#).

- **Schema.org (2 Junio 2011)**

Google, Yahoo! y Microsoft unen fuerzas, al igual que ocurriera con los sitemaps, para fundar la iniciativa Schema.org [\[29\]](#), que tiene como objetivo crear y dar soporte a un conjunto de esquemas que sirvan para consolidar el uso de los datos estructurados en las páginas web. A finales de 2011, Yandex se unió al proyecto. Algunos de estos esquemas existían previamente, otros se han creado como parte de esta alianza.

Los motores de búsqueda son capaces de identificar patrones (palabras clave) dentro de una página web, pero no son capaces de asociar un significado a dichas páginas. El marcado de datos estructurados nos permite dotar de un contexto y un significado al contenido de una página web. Se trata por tanto de uno de los primeros pasos hacia la búsqueda semántica y hacia la búsqueda basada en entidades, a las que Google daría continuidad en 2012 con la introducción del Knowledge Graph y en 2013 con Hummingbird [\[30\]](#) [\[31\]](#).

- **Query Encryption (18 Octubre 2011)**

Google anuncia que, por motivos de seguridad y privacidad, comenzará a hacer uso del protocolo de encriptación SSL en las consultas [\[32\]](#). En un principio su aplicación se restringía a aquellos usuarios que estuvieran registrados en su cuenta de Google en el momento de realizar la búsqueda, que por defecto eran redirigidos a la versión HTTPS de Google. Posteriormente, durante el 2013, su aplicación se extendió al resto de usuarios [\[33\]](#). Desde el punto de vista de estos supone una importante mejora en la forma en la que se protegen sus datos.

Desde el punto de vista del SEO supone un duro golpe. Hablaremos más de las consecuencias de esta actualización durante el proceso de investigación de palabras clave, en la Sección 3.2.1.2.

- **Knowledge Graph (16 Mayo 2012)**

En su presentación, Google empleó la frase “things, not strings” [\[34\]](#). Y es que hasta ese momento la búsqueda se basaba básicamente en asociar keywords y consultas. Para el motor de búsqueda las palabras que componen una consulta son solo eso, palabras.

Con la introducción de Knowledge Graph, Google ha elaborado un modelo inteligente que comprende entidades del mundo real y su relación entre ellas. De esta forma se incorpora una nueva forma de interpretar y de mostrar la información mediante la incorporación de los “Knowledge panels” y las “Knowledge cards” dentro de las SERPs, paneles en los que Google ofrece información instantánea acerca de personas, lugares o cosas sobre las que tiene conocimiento. En el momento de su lanzamiento, en el año 2012, Knowledge Graph ofrecía información sobre ciudades, personajes famosos, edificios, equipos deportivos, películas, características geográficas, objetos celestiales, música, arte, ... contando con más de 500 millones de entidades y más 3.5 billones de hechos y relaciones entre dichas entidades.

Knowledge Graph amplía notablemente el uso de la búsqueda semántica y de la búsqueda basada en entidades que se daba los primeros pasos con la creación de Schema.org y su conjunto de esquemas para el marcado de datos estructurados [\[30\]](#) [\[31\]](#).

- **HTTPS/SSL Update (6 Agosto 2014)**

Tras meses de especulación Google anunció que el uso del protocolo HTTPS pasaría a formar parte de los factores de clasificación [\[35\]](#). Al comienzo su incidencia fue menor, afectando a un 1% de las consultas. En palabras de la propia Google la influencia de este factor se verá incrementada a lo largo del tiempo.

1.3.1.2. Actualizaciones mayores

Se trata de actualizaciones que sacudieron el panorama del SEO debido, en parte, a que modificaron la forma en la que trataban algunos los pilares fundamentales del SEO, como son el contenido (Panda) y los enlaces (Penguin), incluyendo penalizaciones por su uso inadecuado. También por el hecho de combatir y penalizar, en mayor o menor medida, algunos comportamientos inadecuados (piratería, anuncios, spam...).

A continuación, repasamos algunas de estas actualizaciones, haciendo hincapié en Panda y Penguin, posiblemente las más famosas, que se examinarán con mayor detalle.

- **Freshness update (3 Noviembre 2011)**

Google anuncia que su nueva actualización en el algoritmo comenzará a recompensar la ‘frescura’ de la información [\[36\]](#). Esta actualización afecta principalmente a aquellos resultados cuya vigencia puede estar afectada por el tiempo, con el fin de proporcionar los resultados que están más actualizados.

Resulta evidente que no todas las consultas de búsqueda tienen la misma vigencia, y por ello el algoritmo es capaz de aplicar diferentes niveles de ‘frescura’. Desde eventos recientes que requieren una actualización casi inmediata, pasando por eventos recurrentes como

pueden ser los Juegos Olímpicos, hasta elementos que raramente cambian como puede ser una receta.

Esta actualización tuvo un impacto notable, de un 6-10% dependiendo del idioma y del dominio de búsqueda.

- **Panda (24 Febrero 2011)**

Panda se centra en analizar el contenido de las páginas web, determinar la calidad del mismo y en función de dicha calidad, junto con otros múltiples factores, determinar su ranking. En palabras de Google “Esta actualización está diseñada para reducir el ranking de los sitios de baja calidad, es decir, sitios con bajo valor añadido para los usuarios, que copian contenido de otros sitios web, o que no son de utilidad. Al mismo tiempo, proporcionará un mejor ranking para los sitios web de alta calidad, sitios con contenido original como investigaciones, informes detallados, análisis reflexivo...” [\[37\]](#). Esta actualización supone uno de los mayores cambios en el panorama del SEO. En concreto, y en palabras de la propia Google, “afectó a un 12% de las consultas” [\[37\]](#).

En un principio Panda se centró únicamente en EE. UU. Esto cambió el 12 de Agosto del 2011 cuando se lanzó de forma global, lo que afectó a un 6-9% de las consultas internacionales. En sus inicios se trataba más bien de un filtro que Google aplicaba periódicamente con cada una de sus actualizaciones. En Enero de 2016 Panda pasó a formar parte del núcleo del algoritmo de Google [\[38\]](#). Esto no quiere decir que Panda se aplique a los resultados de búsqueda en tiempo real, pero sí implica que la detección del contenido, la aplicación de la penalizaciones y recuperación de las mismas han visto recortados sus plazos. Hasta la fecha Google ha llevado a cabo un total de 30 actualizaciones del algoritmo Panda.

Áreas objetivo de Panda

Históricamente Google ha ofrecido muy poca información acerca del funcionamiento del algoritmo a la hora de determinar la calidad de un sitio web. A continuación, tenemos una relación de las principales áreas objetivo de Panda [\[39\]](#):

- *Thin content/Contenido pobre*

Se trata de páginas con muy poco contenido o con contenido de poco valor para el usuario, como pueden ser los perfiles de usuario en un foro, o los comercios online con miles de páginas de producto, pero muy poco contenido en ellas.

- *Contenido no original/Plagio*

Páginas web cuyo contenido ha sido obtenido mediante técnicas de scraping (copiar y pegar), o páginas que han sido modificadas levemente. Todas ellas son fácilmente detectables por Google. Un sitio web con un número reducido de este tipo de páginas puede sufrir el impacto de Panda.

- *Contenido no diferenciado*

Es posible que el contenido sea original, pero no es suficiente. Si un sitio web trata sobre temas sobre los que ya se ha escrito miles de veces, es posible que Google considere que ese sitio web no aporta nada nuevo y este no aparezca en el índice.

- *Contenido duplicado*

Se trata de contenido duplicado dentro de un mismo sitio web. En el caso de que se produzcan este tipo de errores lo mejor sería eliminar dichas páginas duplicadas. Si por algún motivo esto no fuera posible será necesario emplear la redirección 301 o la etiqueta *rel="canonical"*.

- *Contenido de baja calidad*

Se trata de contenido inexacto o mal redactado. Suele ser difícil de detectar, algunos de los indicadores son la falta de variedad gramática o los errores de ortografía.

- *Thin slicing (Rebanado fino)*

Se trata de páginas que están diseñadas con el único objetivo de ganar tráfico proveniente de las búsquedas. Para ello los editores del sitio web crean una página para cada consulta de búsqueda que el usuario pueda realizar. Por ejemplo, creando páginas con títulos como “software engineering schools”, “software engineering school”, “software engineering colleges”, “software engineering university”, ...

- *Contenido generado a partir de bases de datos*

Esta práctica por sí misma no es mala. El problema es cuando se realiza a gran escala y como resultado se obtiene un gran número de páginas web con muy poco contenido o de escasa calidad. Demasiadas páginas de este tipo pueden provocar que Panda actúe sobre el sitio web penalizándolo en los rankings.

Solución

La solución pasa por llevar a cabo un análisis exhaustivo del sitio web con el objetivo de encontrar las páginas con contenido que pueda verse afectado por la penalización de Panda. Se trata de un proceso en el que es importante contar con una visión crítica del sitio web, algo que no siempre es fácil. Por ello se recomienda, siempre que se posible, contar con una opinión externa, que no esté involucrada en el sitio web afectado.

Una vez se han obtenido los datos existen múltiples soluciones al problema: mejorar el contenido afectado, emplear la etiqueta *noindex* para eliminar del índice páginas afectadas, eliminar las páginas y emplear la redirección 301 para redirigir a los usuarios a otras secciones del sitio web, emplear la herramienta para la eliminación de URL proporcionada por Google...

Una vez los cambios hayan sido realizados, Google tiene que volver a rastrear el sitio web para observar las modificaciones. Antes de la actualización de Enero de 2016, este proceso podría llevar varios meses, hasta que Google determinará que los cambios son suficientes

para que el sitio web supere la penalización. Tras esta actualización, en la que Panda pasa a formar parte del núcleo del algoritmo, este proceso se ha visto agilizado.

- **Top heavy update (19 Enero 2012)**

Se trata de un algoritmo encargado de analizar el diseño de la página y penalizar a aquellas páginas que hacen un uso abusivo de los anuncios, especialmente en la parte superior de las páginas, lo primero que los usuarios ven. Afecta a todas las plataformas de anuncios, incluida AdSense, la plataforma de anuncios de Google. Para evitar las penalizaciones basta con evitar hacer un uso excesivo de los anuncios en la parte superior de las páginas [\[40\]](#).

- **Penguin (24 Abril 2012)**

Los backlinks, es decir, los enlaces recibidos desde otras páginas web son uno de los principales factores que los motores de búsqueda tienen en cuenta a la hora de establecer la posición de una página web. La importancia de su número total es relativa, lo que realmente importa es la calidad y la autoridad de los sitios web que realizan dichos enlaces. El hecho de que un sitio web de calidad y con autoridad contrastada (Wikipedia, The Washington Post...) enlace tu contenido es un claro indicativo de que el contenido también lo es. Pero también tenemos la situación diametralmente opuesta, si un sitio web recibe enlaces de sitios web considerados de baja calidad, Google considerará que ese contenido también lo es.

Google Penguin tiene como objetivo identificar y penalizar aquellos sitios web con un perfil de enlaces poco natural. Se considera que estos sitios web tratan de modificar en su favor los resultados de búsqueda empleando técnicas de obtención de enlaces de carácter manipulativo. Al igual que sucedió con la salida de Panda, Penguin provocó un terremoto en el panorama del SEO, cambiando para siempre las técnicas de obtención de enlaces. Aunque la incidencia fue menor que en el caso de Panda, afectando al 3,1% de las consultas en inglés [\[41\]](#).

Áreas objetivo de Penguin

Como ya hemos comentado, Penguin persigue los enlaces sospechosos/malos/ de baja calidad. Existen numerosos tipos de enlaces de mala calidad. A continuación, veremos una relación de los tipos de enlace sobre los que Penguin pone el foco [\[42\]](#).

- *Links que provienen de sitios de baja calidad.*

Uno de los aspectos más relevantes de un enlace es la calidad y la autoridad del sitio web que realiza dicho enlace. Mientras que los enlaces de calidad pueden ayudar a mejorar significativamente la posición de una página, los enlaces de provenientes de sitios web de baja calidad, normalmente relacionados con el spam, tienen el efecto contrario y pueden llevar a una penalización.

- *Links que proceden de sitios web cuya temática sea irrelevante*

Resulta sospechoso que p. ej. un sitio web como el que estamos tratando de este proyecto, con una temática tan particular y especializada, reciba enlaces desde un

blog de moda, deportes... Puede suponer un indicativo de que el sitio web está empleando técnicas de obtención de enlaces de carácter manipulativo.

- *Directorios de artículos*

Desde un principio Google puso en el objetivo aquellos enlaces que provienen de directorios de artículos. Aunque estos enlaces no siempre tienen un carácter manipulativo, Google considera que aquellos sitios web que obtienen enlaces de los directorios de artículos tienden a ser de mala calidad.

- *Directorios web*

Un directorio web es una lista de sitios web, generalmente dividido en diferentes categorías en función de su temática. Suponían una forma fácil de obtener un backlink de relativa calidad, en algunos casos previo pago. Es una “técnica” ya en desuso, pero en su momento existieron un gran número de directorios web y, salvo los más conocidos (Yahoo! Directory, DMOZ, Business.com), el resto pueden suponer una penalización. De hecho, en estos momentos solo Business.com sigue operando.

- *Links de pago*

Google especifica en las Directrices para Webmasters que los enlaces de pago deben llevar la etiqueta “rel=nofollow”. De esta forma estos enlaces no serán tenidos en cuenta por Google a la hora de posicionar dicha página web.

- *Links con un anchor text sobre optimizado*

La existencia de numerosos enlaces con exactamente un mismo texto de enlace, rico en palabras clave, es indicativo de que dicho sitio web está empleando técnicas de spam.

- *Enlaces de países poco relevantes para el negocio*

Si un sitio web opera únicamente en una determinada región/idioma y recibe enlaces desde otras regiones/idiomas puede ser tomado como un comportamiento sospechoso y un posible signo de spam.

- *Spam en la sección de comentarios*

Se trata de uno de los principales y más extendidos métodos de spam. El hecho de que los enlaces de un sitio web aparezcan con frecuencia en la sección de comentarios de blogs y foros de terceros, será tomado como un signo de spam y penalizado por Penguin.

Penalizaciones y su solución

Hasta finales de 2016 las penalizaciones provocadas por Penguin eran fáciles de detectar, puesto que se corresponden con las fechas de actualización del algoritmo. Por desgracia también eran las únicas fechas en las que un sitio web podía recuperarse. A partir de finales de 2016 Penguin pasa a formar parte del núcleo del algoritmo principal [\[43\]](#) y comienza a

actuar en tiempo real, lo que reduce notablemente los plazos de detección, penalización y recuperación, que anteriormente podían llegar al año.

La solución pasa por llevar a cabo una auditoría de los enlaces que recibe el sitio web. En el caso de que se detecten enlaces perjudiciales la situación ideal pasaría por contactar con los Webmasters de los sitios web en cuestión, y solicitar la retirada de dichos enlaces. Si esto no fuera posible Google, a través de GSC, pone a disposición de los webmasters la herramienta Google Disavow [\[44\]](#), que nos permite desautorizar los enlaces sospechosos.

- DMCA Penalty/ Pirate (10 Agosto 2012)

Esta actualización del algoritmo tiene como objetivo la aplicación de penalizaciones a aquellos sitios web que reciben un número elevado de peticiones de retirada de contenido (formulario en [\[45\]](#)) por infracción de Copyright/DMCA [\[46\]](#). La única forma de evitar este tipo de penalizaciones, como es obvio, es publicar únicamente tu propio contenido. En el caso de publicar contenido de terceros será necesario contar con su consentimiento.

- Exact-match domain (EMD) (27 Septiembre 2012)

Hasta la puesta en marcha de esta actualización, los dominios que coincidían exactamente con la consulta realizada (p. ej. <http://www.zapatos-azules.com> y *zapatos azules*) contaban con una clara ventaja a la hora de establecer su posición en el ranking para dicha consulta. Con esta actualización Google cambia la forma en la que valora los exact-match domain, reduciendo e incluso eliminando dicha ventaja.

- “Payday loan” update (11 Junio 2013)

Lanzado en Junio de 2013, este algoritmo tiene como objetivo aquellos sitios web y consultas de búsqueda que pertenecen a mercados con un índice de spam particularmente alto [\[47\]](#). Google pone como ejemplo los préstamos (payday loan), los casinos y la pornografía como mercados objetivo. El algoritmo afectó a un 0,3% de las consultas en Estados Unidos, llegando a afectar un 4% de las consultas en Turquía, donde el uso de técnicas de spam es más alto.

- Hummingbird (20 Agosto 2013)

Hummingbird no es una actualización al uso, se trata de una reescritura de la plataforma de búsqueda de Google con el fin de hacerla más flexible y adaptable al futuro. Por ello se incluyó un mayor soporte para la búsqueda semántica y para aquellas consultas que emplean un lenguaje conversacional, debido entre otras cosas al aumento de las búsquedas desde dispositivos móviles, que hace un mayor uso de este tipo de lenguaje [\[48\]](#).

Con Hummingbird cambia la forma en la que se interpretan las consultas. Hasta ese momento la importancia de las keywords que conforman la consulta era total. Hummingbird desarrolla el concepto de búsqueda semántica, en el que se trata de determinar la intención del usuario, el contexto de la consulta, más allá de las palabras que la componen. De esta forma puede proporcionar resultados que se ajustan mucho mejor a lo que el usuario busca. Evidentemente las keywords que forman la consulta siguen siendo importantes, pero Hummingbird da un mayor poder al significado completo de la consulta. Así mismo,

Hummingbird incluye una optimización en el uso de sinónimos. De esta forma Google ya no muestra en sus SERPs una lista de resultados que coincidan exactamente con las keywords de la consulta. Ahora se muestran resultados relacionados con el tema, que no tienen porque incluir necesariamente las keywords empleadas en la consulta.

Hablaremos más del funcionamiento de este algoritmo durante el proceso de investigación de palabras clave, en la Sección 3.2.1.2.

- **Mobile update (22 Abril 2015)**

El principal motivo de esta actualización reside en un porcentaje cada vez mayor de búsquedas realizadas desde dispositivos móviles (smartphones), superando desde el año 2015 a las búsquedas desde escritorio [\[49\]](#). Con estos datos, y dado que la experiencia de usuario en un smartphone es muy diferente de la que podemos tener a la hora de visitar una página en un dispositivo de escritorio, Google decide potenciar la optimización para dispositivos móviles, pasando a formar parte de los factores que afectan a la posición de una página web [\[50\]](#). Aquellas páginas que estén optimizadas verán potenciada su posición, mientras que las que no estén podrían sufrir pérdidas. Esto solo afecta a las búsquedas realizadas desde smartphones, por tanto, las búsquedas realizadas desde dispositivos de escritorio no se verán afectadas.

- **RankBrain (26 Octubre 2015)**

En octubre de 2015 Google anunció que RankBrain, su sistema de aprendizaje automático, llevaba meses formando parte del algoritmo. En el comienzo de su aplicación afectaba a alrededor de un 15 % de las consultas, en estos momentos y tras demostrar su fiabilidad su aplicación es universal. En palabras de Google se trata del tercer factor más influyente para la clasificación de una página web dentro del ranking [\[51\]](#).

En esencia la función de RankBrain es ayudar a Google a entender mejor el significado que hay detrás de cada consulta y de esta forma ofrecer los resultados que mejor se ajusten a dicha consulta. Para ello hace uso de las técnicas de aprendizaje automático. Por ejemplo, si RankBrain se encuentra con una consulta con palabras o frases con las que no está familiarizado o que resulten ambiguas, la máquina puede tratar de adivinar, en base a su actual conocimiento, que palabras o frases tienen un significado similar y en base a ello filtrar los resultados, haciendo de esta forma más efectivo el hecho de tratar con consultas desconocidas o ambiguas.

RankBrain supone el tercer paso en el afán de Google por lograr entender el significado de las consultas que realizan los usuarios, el contexto detrás de cada palabra, con el fin de devolver respuestas mucho más precisas. Anteriormente, Google introdujo en 2012 el Knowledge Graph, y con él el concepto de Búsqueda semántica y de búsqueda basada en entidades. Más tarde, en 2013, Google lanzó Hummingbird, que asentó definitivamente la búsqueda semántica haciéndola parte del núcleo del algoritmo de búsqueda.

1.4. Objetivos

El principal objetivo de este proyecto es optimizar el posicionamiento en los motores de búsqueda de las páginas del sitio web del Multiscale Materials Modeling Group (<https://www.ele.uva.es/~mmm/>), y en especial de la página más relevante del sitio web, la página principal. Dada la posición de mercado dominante que Google posee, centraremos nuestros esfuerzos en optimizar nuestras páginas web para dicho buscador, sin perder de vista otros motores de búsqueda.

De esta forma podremos cumplir con otro de nuestros objetivos, que no es otro que tratar de establecer que influencia tiene cada una de las modificaciones llevadas a cabo durante el proceso de optimización sobre la posición de las páginas web en los resultados de búsqueda. También compararemos la influencia de dichas modificaciones entre los distintos motores de búsqueda.

1.5. Metodología

En primer lugar, determinaremos cuales son los factores SEO que los motores de búsqueda tienen en consideración a la hora de establecer la posición de las páginas web en los resultados de búsqueda.

Antes de comenzar el proceso de optimización del sitio web, analizaremos el estado previo del mismo para tratar conocer mejor su estructura, su estado de indexación, ... y tratar de detectar elementos o situaciones que son susceptibles de ser mejoradas.

Posteriormente, aplicaremos las modificaciones que resulten pertinentes para adaptar nuestro sitio web a los requerimientos impuestos por los distintos factores y que nos permitan solucionar los problemas descubiertos durante el análisis previo del sitio web.

Finalmente, se llevará a cabo el seguimiento de las posiciones que ocupan nuestras páginas web para un conjunto de palabras clave, representativas del contenido de sitio web. Este seguimiento nos permitirá conocer la influencia que cada una de las modificaciones tiene sobre la posición de la página web en cuestión.

1.6. Estructura de la memoria

En el Capítulo 1 se han establecido las bases de este proyecto. Hemos definido conceptos básicos, se ha explicado brevemente el funcionamiento de los motores de búsqueda, se ha analizado el estado del mercado de los buscadores, y se ha repasado el historial de actualizaciones del dominador histórico del mercado, Google.

En el Capítulo 2 estableceremos una relación de los principales Factores SEO que los motores de búsqueda tienen en consideración a la hora de establecer la posición de las páginas web en los resultados de búsqueda.

El Capítulo 3 estará centrado en la optimización del sitio web del que es objeto este proyecto. Para ello aplicaremos las modificaciones que resulten pertinentes con el fin de adaptar las páginas del sitio a los factores anteriormente descriptos.

En el Capítulo 4 trataremos el proceso de seguimiento y analizaremos los resultados obtenidos en el mismo.

Finalmente, en el Capítulo 5, expondremos las conclusiones derivadas del análisis de los resultados y de la elaboración del propio proyecto.

2. Factores SEO

Son el secreto mejor guardado por parte de los motores de búsqueda. Google cuenta con más de 200 factores que afectan a la posición en el ranking de una página web, tanto de forma positiva como negativa [52]. Muchos de estos factores cuentan a su vez con hasta 50 variaciones lo que podría elevar el número hasta los 10.000 factores, en contraposición a los 1000 factores de Bing [53].

Algunos de estos factores son de sobra conocidos, básicamente porque la propia Google ha reconocido su utilización, otros en cambio se “presuponen”. Lo que todos ellos tienen en común es que se desconoce su incidencia exacta a la hora de posicionar una página web.

Elaboraremos nuestra propia lista de factores SEO tomando como referencia varias fuentes: distintas guías SEO como la de Moz [54] o la propia SEO Starter Guide de Google [55], el historial de actualizaciones del algoritmo que vimos en la sección 1.3, estudios llevados a cabo por empresas especializadas en SEO...

- Search Engine Land

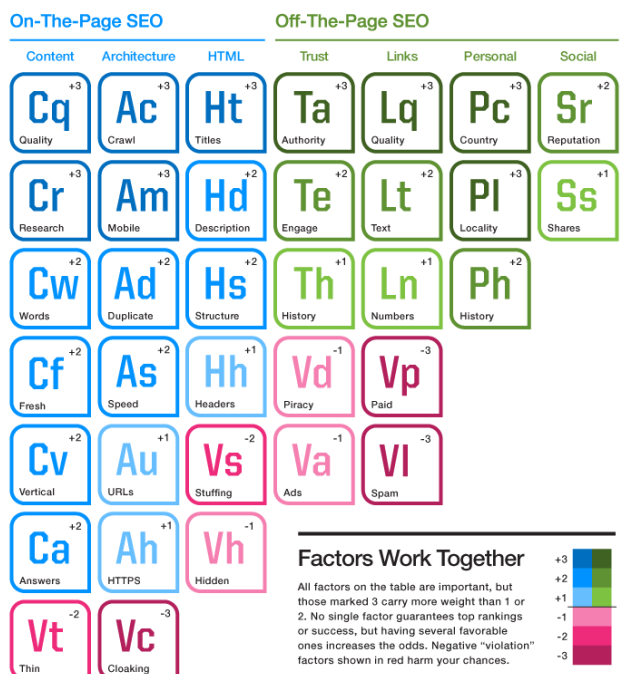
Search Engine Land (SEL) es una publicación especializada en la industria del marketing en buscadores. Lleva a cabo una cobertura de las últimas novedades en el sector, como pueden ser cambios en el algoritmo o la incorporación de nuevas tendencias al mercado. Además, publica artículos de opinión firmados por expertos en marketing digital, en los que se lanzan consejos prácticos para lograr llevar a cabo campañas de marketing de forma exitosa. También organiza conferencias y seminarios especializados en la materia.

Centrándonos en el tema que nos concierne, tomaremos una de las producciones más exitosas de SEL, “The Periodic Table of SEO Success Factors”. Cada dos años, SEL lleva a cabo una encuesta entre sus usuarios en la que pregunta acerca de determinados factores y su importancia relativa a la hora de establecer la posición de una página web.

Como explican desde SEL, la encuesta no trata de listar los más de 200 factores que Google emplea a la hora de establecer el ranking, sino que más bien trata de establecer los factores generales del SEO, sin entrar en el detalle de aspectos como p. ej. si la etiqueta de título de HTML tiene una longitud adecuada...

The Periodic Table of SEO Success Factors

Search engine optimization (SEO) seems like alchemy to the uninitiated. But there's a science to it. Below are some important "ranking factors" and best practices that can lead to success with both search engines and searchers.



WRITTEN BY: Search Engine Land CREATED BY: COLUMN FIVE LEARN MORE: <http://selind.com/seotable> © 2017 Third Door Media

Figura 8. The Periodic Table of SEO Success Factors. 2017.

Como vemos en la Figura 8, correspondiente con la versión del 2017 de la Tabla periódica de los factores SEO [\[56\]](#), los factores se dividen en dos apartados, dependiendo de si estos son aspectos internos de la propia página o bien son factores externos. A su vez cada uno de los apartados cuenta con secciones que nos permiten dividir los factores en función de sus características. Cada sección cuenta con una serie de factores algunos positivos, otros negativos, a los que se le asigna una puntuación acorde con su importancia relativa. SEL lleva a cabo una guía denominada “Search Engine Land’s Guide To SEO” en la que explican en profundidad cada uno de los factores [\[57\]](#).

- MOZ

MOZ es una empresa proveedora de servicios de software especializada en marketing en buscadores. Cuenta con un conjunto de herramientas de marketing analítico cuya utilización requiere suscripción. Además, cuenta con un blog en el que, al igual que SEL, se publican artículos con las últimas novedades del sector, guías para usuarios, casos de estudio, ...

Cada dos años, MOZ lleva a cabo un estudio de los factores más influyentes a la hora posicionar una página en los motores de búsqueda [\[58\]](#). El estudio cuenta con dos partes, por un lado, Moz realiza una encuesta a 150 expertos SEO a los que les pide que evalúen la importancia de determinados factores SEO, divididos en grupos dependiendo de su área de influencia y con el fin de cubrir todos los aspectos generales del SEO. En este caso se trata de una encuesta más exhaustiva que la que realiza SEL, que como vimos anteriormente trata de establecer unas pautas generales.

Por otro lado, Moz lleva a cabo un análisis de los resultados de búsqueda de “17.600 keywords” en Google.com (US). Con estos datos Moz realiza un análisis estadístico en el que trata de establecer correlaciones entre la aparición de determinados factores en las páginas web y un mejor posicionamiento de dichas páginas en las SERPs. Cabe recordar que se trata de correlaciones, no de causalidades. Es decir, por el simple hecho de que las páginas web que tienen una determinada característica cuenten con una posición más alta, no implica que necesariamente dicha característica sea la causante dicha posición más alta.

Se trata de un estudio demasiado exhaustivo y complejo para ser incluido en estas páginas. Trataremos de combinar la simplicidad y la estructura de los datos obtenidos de la Tabla Periódica del SEO de Search Engine Land con datos del estudio de Moz, con el fin de que estos últimos nos den una mayor profundidad y lograr una versión intermedia.

2.1. Lista de Factores SEO

Existen numerosos factores que influyen, en mayor o menor medida, en la posición que ocupa una página web en la página de resultados para una determinada consulta. Como hemos mencionado anteriormente, Google emplea más de 200 factores para establecer dicha posición [\[52\]](#), y en muchos casos no se conoce a ciencia cierta su identidad ni la incidencia que estos tienen en el posicionamiento de una página web.

El objetivo de esta sección es elaborar una lista con los principales factores que los motores de búsqueda tienen en consideración a la hora de establecer la posición de una página web en las SERPs. Por tanto, en este proyecto no cubriremos todos y cada uno de los factores, aplicaremos un enfoque

más general dividiendo los factores en áreas de influencia, cada una de ellas con una serie de elementos a tener en cuenta. En concreto dividiremos los factores SEO en tres grandes categorías en función de su tipología: On-Page, Off-Page y Negativos.

Una vez conocidos estos factores la lista nos servirá de guía para llevar a cabo las modificaciones que permitan optimizar la posición de las páginas de nuestro sitio web. Por ello, en este apartado nos limitaremos a realizar una pequeña descripción de cada uno de los factores. Posteriormente, durante el proceso de optimización del sitio web, se profundizará en cada uno de los factores

Antes de comenzar con la lista cabe señalar que, aun teniendo diferentes grados influencia, ninguno de los factores que veremos a continuación logrará grandes resultados funcionando por sí solo.

2.1.1. Factores SEO On-Page

Esta categoría recoge aquellos factores que están bajo el control del webmaster del sitio web. Se trata por tanto de aspectos relativos al contenido publicado, a la arquitectura y elementos técnicos del sitio web y al propio código HTML de cada una de las páginas [\[56\]](#) [\[58\]](#).

2.1.1.1. Contenido

El contenido es el activo más importante de cualquier sitio web, y sienta las bases del éxito de una campaña SEO. Sin un contenido de calidad resultará difícil que el resto de cambios involucrados en el proceso obtengan los resultados esperados. A continuación, se establece una relación de factores relacionados con el contenido que influyen, en mayor o menor, medida en la posición de un sitio/página web dentro del ranking:

- **Contenido de calidad**

Cabe hacerse las siguientes preguntas, ¿ofrece a sus usuarios contenido único, diferente, útil, que no pueda encontrar en ningún otro lugar? ¿algo que les haga permanecer en la página web y que les haga volver en el futuro?

- **Contenido fresco**

Los motores de búsqueda tienen predilección por los nuevos contenidos y tratan de indexarlos con el fin de ofrecer a los usuarios los resultados más relevantes y recientes para cada una de las consultas. Pero, resulta evidente que no todas las consultas tienen las mismas necesidades de actualización, por ello Google cuenta con algoritmos capaces de detectar qué consultas necesitan una actualización para incluir nuevos contenidos.

- **Investigación de palabras clave**

Se trata de un proceso fundamental en el que se trata de descubrir las palabras que los usuarios emplean a la hora de buscar el contenido del sitio web. De poco sirve contar con contenido de calidad si este no se ajusta a las palabras clave que emplean los usuarios en sus consultas. Por tanto, resultará conveniente emplear dichas palabras clave en la elaboración del contenido, de forma que este pueda dar una respuesta más efectiva a dichas consultas. Se llevará a cabo una descripción mucho más detallada de todos los aspectos involucrados en el proceso en la sección 3.2.1.2.

- **Uso de las palabras clave en el contenido**

Tras el proceso de investigación de palabras clave llega el momento de analizar el uso que se hace de dichas palabras en el sitio web, tanto en su estructura como en su contenido. Existen métricas que miden la densidad de palabras clave, pero no existe una cifra ideal que garantice que la página vaya a mejorar su posicionamiento. Se trata más bien de emplear el sentido común y un lenguaje natural.

- **Contenido vertical**

Cuando hablamos de contenido vertical nos referimos al uso de imágenes, videos, noticias, negocios locales, ... Las páginas de resultados comúnmente incluyen “resultados verticales” que pueden dar a la página web una posición que de otra forma no conseguiría.

- **Gestión del contenido duplicado**

Existen dos tipos de contenido duplicado, ambos perjudiciales para el SEO. En este caso hacemos referencia a la gestión del contenido duplicado “externo”, es decir, a la publicación por nuestra parte de contenido que se encuentra en otros sitios web. Si el motor de búsqueda descubre que el contenido ha sido copiado, el sitio web puede verse afectado por penalizaciones.

2.1.1.2. Arquitectura y aspectos técnicos

Una arquitectura óptima puede hacer que el proyecto SEO se vea agilizado en algunos de sus procesos (rastreo, indexación...), en cambio, una mala arquitectura puede paralizarlos. Se incluyen además aspectos técnicos que pueden mejorar el posicionamiento del sitio web.

- **Arquitectura óptima**

El objetivo es lograr una arquitectura que sea accesible y fácil de utilizar tanto para los motores de búsqueda como para los humanos. Por tanto, se tratará de evitar hacer uso de elementos que entorpezcan la tarea de rastreo e indexación llevada a cabo por los motores de búsqueda y haremos uso de una estructura de enlaces que facilite la navegación y la jerarquización del contenido.

- **Formato de las URLs**

El formato de las URLs no es, en sí mismo, un factor que afecte directamente a la posición de una página web. Pero sí que lo hace de forma directa al estar relacionado con otros factores como la usabilidad y la experiencia de usuario. Se trata de aspectos como las palabras empleadas, su tamaño, ...

- **Optimización para dispositivos móviles**

Desde el año 2015 el número de búsquedas desde dispositivos móviles es mayor que él búsquedas en escritorio [\[49\]](#). Y por ese motivo desde los motores de búsqueda se potencia la optimización para dispositivos móviles ofreciendo la posibilidad de que aquellas páginas que estén optimizadas puedan lograr un mejor posicionamiento en las búsquedas móviles frente a aquellas páginas que no lo están.

- **Tiempo de carga**

Desde el año 2010, las páginas con un tiempo de carga menor tienen una pequeña ventaja frente a las páginas más lentas. Se trata de un factor menor que afecta, en palabras de Google, al 1% de las consultas [\[59\]](#). Aun así, una página web rápida da buena imagen entre los usuarios y puede suponer la diferencia entre una visita o un rebote hacia la página de resultados.

- **HTTPS/SSL**

Desde el año 2014 Google recompensa el uso de HTTPS/SSL con una pequeña mejora en el posicionamiento. En sus inicios, al igual que el tiempo de carga, se trataba de un factor menor, que por sí solo no garantiza una mejora sustancial. En palabras de la propia Google la influencia de este factor se verá incrementada a lo largo del tiempo.

- **Gestión del contenido duplicado**

El objetivo es que exista una única versión de cada página de un sitio web. Cuando tenemos múltiples versiones de una misma página web el motor de búsqueda no mostrará todas ellas, tomará la que considere canónica y el resto quedarán ocultas. Además, si cada una de las versiones obtiene enlaces, la autoridad que los usuarios proporcionan al enlazar el contenido se verá diluida entre el número de copias, y no se verá reflejada la importancia real que los usuarios dieron a dicho contenido.

2.1.1.3. Código HTML

Los motores de búsqueda tienen en cuenta algunos de los elementos albergados en las etiquetas HTML a la hora de determinar el posicionamiento de una página. Se trata también de una cuestión de usabilidad, pues en muchos casos es el primer punto de contacto con el usuario. Veamos cuales son los más importantes:

- **Etiqueta de título**

Se trata del elemento más importante dentro del código HTML, pues los motores de búsqueda lo emplean, junto con otros elementos, para determinar la temática de la página. Por tanto, los títulos deberán contar con una serie de características: ser únicos, descriptivos, tamaño, ... Además, se trata también de una cuestión de usabilidad que afecta a la experiencia de usuario, pues las SERPs suponen el primer punto de contacto con los usuarios, y si el título no responde a sus expectativas pueden descartar visitar la página.

- **Etiqueta de descripción**

Nos permite proporcionar una descripción del contenido, que será visible en las SERPs. No se trata de un elemento que mejore directamente el posicionamiento de una página, se trata nuevamente de un elemento estético que nos permite mejorar la experiencia de usuario, pues junto con la etiqueta de título supone el primer contacto de los usuarios con la página y puede marcar la diferencia a la hora de obtener visitas. Existen una serie de aspectos con los que deben contar: únicas, descriptivas, tamaño...

- **Etiquetas de encabezado**

Las etiquetas de encabezado(<h1> - <h6>) se emplean para dotar al contenido de una estructura y para identificar las principales secciones, tanto para los usuarios como para los motores de búsqueda, que emplean dichas etiquetas para tratar de averiguar la temática de la página web.

- **Anchor text/Texto de anclaje**

El anchor text o “texto de anclaje” es considerado por los motores de búsqueda como la forma en que la persona que realiza el enlace describe el contenido enlazado.

- **Nombre de archivo y atributo *alt* de las imágenes**

El nombre de archivo y el atributo *alt*, presente en las imágenes, nos permite dotar a las imágenes con una descripción de su contenido lo que permite a los motores de búsqueda establecer mejor el contexto de la imagen y, por tanto, indexar las imágenes de forma adecuada.

- **Datos estructurados**

El uso de un lenguaje de marcado de datos estructurados nos da la posibilidad de ayudar a los motores de búsqueda a entender mejor el significado del contenido de una página, así como su estructura. Además, nos puede permitir obtener un posicionamiento relevante si nuestro contenido aparece en los paneles del Knowledge Graph. También tiene un componente estético pues el marcado de los ratings o valoraciones permite la aparición de los denominados “rich snippets”, que dan un valor añadido a los resultados en las SERPs. Lo mismo ocurre con el marcado de los Breadcrumbs, que es el que emplearemos en este proyecto pues es el que más se ajusta a nuestras necesidades.

- **Open Graph y Twitter Cards**

Se trata de un conjunto de etiquetas que nos permiten marcar ciertos elementos de un sitio web con el fin de que estos sean interpretados correctamente por las redes sociales como Facebook (Open Graph) y Twitter (Twitter Cards).

2.1.2. Factores SEO Off-Page

Agrupamos aquellos factores sobre los que el webmasters no tienen un control directo. Se trata de aspectos como los enlaces entrantes (Backlinks), las métricas de experiencia de uso, la autoridad/confianza del sitio web percibida por los motores de búsqueda, factores personales de los usuarios...

2.1.2.1. Confianza

Se considera que el grado de confianza que los motores de búsqueda perciben sobre un sitio web juega un importante rol a la hora de determinar el éxito o el fracaso a la hora de posicionar sus páginas.

- **Autoridad**

Se desconoce el método exacto por el cual los motores de búsqueda establecen la autoridad de un dominio, pero es seguro que intervienen múltiples factores como el tipo y número de enlaces que recibe, el número de referencias desde redes sociales, los datos de uso, ...

- **Historial**

Desde el momento en el que los motores de búsqueda comienzan a rastrear el contenido de un sitio web, comienza a registrar su actividad y son capaces de establecer patrones de comportamiento a lo largo del tiempo.

Un sitio web con un buen historial puede hacer más fácil la labor de posicionar su contenido. Así mismo, aquellos sitios web ya establecidos y con un largo historial tendrán facilidades frente a un sitio web de nueva creación, que tardará un tiempo en alcanzar dicho estatus.

En cambio, un sitio web con un historial de penalizaciones y de violaciones de las directrices de uso cargará con su historial durante un tiempo hasta que pueda recuperarse y volver a posicionar sus páginas con normalidad.

- **Datos de uso/métricas**

Las interacciones de los usuarios con las páginas web que aparecen en los resultados de búsqueda son registradas por los motores de búsqueda con el fin de medir el grado de compromiso que muestran los usuarios con el contenido de dichas páginas y la usabilidad de las mismas. Por tanto, pueden significar tanto una muestra de la calidad de las mismas (elevado tiempo de permanencia) como todo lo contrario (los usuarios retornan a la página de resultados inmediatamente). A continuación, establecemos una relación de los datos de uso concretos que emplean los motores de búsqueda:

- *Proporción de clics/Click-through rate (CTR)*: Ratio que resulta de dividir el número de clics que recibe una página entre el número de veces que aparece en las SERPs.
- *Tiempo de permanencia*: Tiempo que los usuarios permanecen en la página.
- *Tasa de rebote/Bounce rate*: Se denomina rebote a aquellas sesiones en las que el usuario visita únicamente una página. La tasa resulta de dividir las sesiones con una sola página visitada entre el número total de sesiones.
- *Interacciones en redes sociales*: El número de comentarios, shares, likes tweets... representan una forma de medir el compromiso del usuario.
- *Visitantes nuevos vs. Visitantes que regresan*: Un buen número de visitantes que regresan al sitio web supone un indicativo de que el contenido es de calidad.

2.1.2.2. Enlaces

Los enlaces son el principal factor externo que los motores de búsqueda tienen en cuenta a la hora de establecer la posición de una página web dentro del ranking. Los motores de búsqueda emplean el análisis de los enlaces a lo largo de la web, comúnmente denominado Link Graph, para determinar la relevancia de los contenidos. A continuación, veremos una relación de factores que influyen en la relevancia que aporta cada enlace, pues no todos los enlaces son iguales:

- **Calidad**

La calidad de un enlace viene marcada por la procedencia del mismo. Los motores no valoran igual un enlace procedente de un sitio web de gran tamaño y reputación como puede ser Wikipedia, The Guardian, Nature..., que un enlace que proviene de la sección de comentarios de un pequeño blog.

- **Relevancia temática del enlace**

También se tiene en cuenta la temática de los sitios web que enlazan el contenido, teniendo mayor valor aquellos enlaces que provienen de sitios web relevantes dentro de la temática del contenido enlazado.

- **Anchor Text/Texto de enlace**

El anchor text o “texto de anclaje” es considerado por los motores de búsqueda como la forma en que la persona que realiza el enlace describe el contenido enlazado. Resulta difícil controlar qué palabras emplean aquellas personas que deciden enlazar tu contenido, pero si existe la oportunidad de influir en los anchor text la aprovecharemos en la medida de lo posible.

La diversidad del anchor text será importante, pues un número elevado de enlaces con el mismo texto puede ser considerado por los motores de búsqueda como spam. Así mismo la sobreutilización de palabras clave dentro de los textos de enlace se considera también una señal de spam.

- **Número**

Como hemos visto los enlaces contribuyen a establecer la relevancia de un contenido a ojos de los motores de búsqueda y, por tanto, es común pensar que cuantos más enlaces tenga una página web mejor será para su posicionamiento. En principio esto es así, pero con matices, pues no solo importa la cantidad sino también la variedad. Los motores de búsqueda dan mucha menos importancia a 100 enlaces procedentes de un sitio web que a 100 enlaces procedentes de 100 sitios web.

- **Posición/contexto de los enlaces**

No tiene el mismo valor un enlace que acompaña al texto de una noticia o entrada de blog, a modo de referencia del mismo, que un enlace ubicado en la sección de comentarios de una web o en una página que albergue un número elevado de enlaces, a modo de listín telefónico.

- **Velocidad de adquisición**

El crecimiento en el número de enlaces debe producirse de manera natural. Los motores de búsqueda consideran sospechoso que existan picos de obtención de enlaces, salvo que se trate de contenido de carácter estacional como pueden ser las navidades, Halloween...

- **Nofollow**

El uso de la etiqueta nofollow en los enlaces hace que todos los factores relativos a los enlaces vistos anteriormente resulten prácticamente inútiles, pues provoca que no se transmita la autoridad/relevancia asociada a dicho enlace.

2.1.2.3. Personalización

Hace unos años todos los usuarios veían los mismos resultados ante una misma consulta de búsqueda. Actualmente los motores de búsqueda personalizan los resultados en función de varios factores que veremos a continuación, de forma que cada usuario obtiene, en mayor o menor medida, una experiencia personalizada.

- **País / Geo-Targeting**

Los motores de búsqueda muestran los resultados que consideran relevantes teniendo en cuenta el país en el que se encuentra el usuario. De esta forma si el sitio web no es considerado relevante para un país en concreto, tendrá menos posibilidades de aparecer en los resultados de ese país.

De ahí la importancia del Top Level Domain (TLD) del sitio web. Mientras que un gTLD (generic Top Level Domain), como .com, .net, .org..., no está asociado a ningún país, un ccTLD (country-code Top Level Domain) como .es, .fr, .uk..., está asociado a un país en concreto y los motores de búsqueda considerarán que su contenido está dirigido al público de dicho país, lo que puede provocar que los resultados en otros países se resientan. Por ello, si se tienen objetivos internacionales el uso de un dominio de tipo ccTLD no resulta recomendable [\[60\]](#).

- **Resultados locales**

El nivel de personalización en base a la localización del usuario profundiza hasta en nivel de la localización exacta, ofreciendo resultados específicos incluso a medida que el usuario se desplaza por la ciudad.

- **Historial**

Los motores de búsqueda personalizan los resultados en función del historial de navegación del usuario, priorizando la posición de aquellos sitios web que el usuario ha visitado con cierta frecuencia. Con ello la primera impresión que recibe un usuario cuando visita un sitio web por primera vez tienen más importancia que nunca.

2.1.2.4. Redes sociales

El uso cada vez más extendido de las redes sociales ha hecho que cobren importancia como factor a la hora de establecer el ranking. La forma en la que se comparte el contenido, mediante “shares”, “likes”, “+1s” ..., sirve de muestra del compromiso de los usuarios con dicho contenido, y por tanto hace que sean considerados para complementar a los enlaces a la hora de establecer la relevancia del contenido.

- **Reputación**

Al igual que a ojos de los motores de búsqueda no todos los enlaces tienen el mismo peso, las cuentas en redes sociales tampoco lo tienen. Por tanto, no tendrá la misma importancia un “share”, “like” ... desde una cuenta con reputación que desde una cuenta cualquiera.

- **Número de shares, likes, +1s...**

De igual forma que con los enlaces, debe existir un equilibrio entre su número y la calidad de los mismos.

2.1.3. Factores negativos

Existen factores que son considerados de forma negativa por los motores de búsqueda. Algunos de ellos, los más perseguidos lo largo de los años a través de las múltiples actualizaciones del algoritmo, forman parte de lo que se conoce como técnicas “Black hat”. Técnicas que buscan manipular la percepción que tiene el motor de búsqueda sobre el sitio web en cuestión con el fin de obtener un mejor posicionamiento.

El uso de las siguientes técnicas puede derivar en penalizaciones que van desde pérdidas en el posicionamiento hasta la total desaparición del índice.

- **Thin content**

Se trata de páginas web que ofrecen contenido con muy poco valor para el usuario. Por ello, los motores de búsqueda no están interesados en su indexación. Un número de páginas de este tipo lo suficientemente elevado puede ser considerado como un indicativo de la baja calidad de un sitio web y provocar las penalizaciones del algoritmo Panda.

- **Cloaking**

Esta técnica consiste en mostrar una versión diferente a los humanos y a los rastreadores de los motores de búsqueda, y de esta forma tratar de manipular los resultados de búsqueda. Se trata de una de las técnicas peor consideradas y más penalizadas.

- **Keyword stuffing**

Se trata de una de las técnicas más antiguas. Consiste en la sobreutilización de las palabras clave con el único fin de tratar de mejorar el posicionamiento, partiendo de la base de que los motores de búsqueda establecen la posición de una página web en función, entre otras cosas, de las palabras clave que aparecen en el contenido de la misma.

- **Contenido oculto**

Sigue la línea del Keyword Stuffing. Consiste en la colocación de contenido no visible por los usuarios, pero sí por los motores de búsqueda, normalmente empleando contenido rico en palabras clave con el mismo color de texto que el fondo de la página. De esta forma el contenido resulta invisible para el usuario, pero no para los motores de búsqueda.

- **Piratería**

La actualización del algoritmo conocida como *Pirate* puso el foco en los sitios web que infligen los derechos de copia(Copyright/DMCA). Por lo tanto, si un sitio web aloja dicho tipo de contenido tarde o temprano será penalizado.

- **Anuncios abusivos**

Los motores de búsqueda penalizan a aquellos sitios web que hacen un uso abusivo de los anuncios, perjudicando la experiencia de usuario.

- **Enlaces de pago**

La compra de enlaces es considerada por Google como un intento de manipular los resultados de búsqueda, y por tanto está perseguida y penalizada, en este caso por el algoritmo *Penguin*.

- **Spam**

Consiste en la creación de enlaces empleando un software automatizado, generalmente con un texto de enlace sobre optimizado. El objetivo está centrado en foros y secciones de comentarios páginas webs. Se trata de una técnica que busca la manipulación de los resultados de búsqueda y por ende resulta altamente perseguida por los motores de búsqueda. En concreto, en el caso de Google, por el algoritmo denominado *Penguin*, encargado de la detección y penalización de dichos comportamientos.

Existe también un tipo de factor negativo que resulta de la no aplicación de algunos de los factores “positivos” relatados en las secciones previas, como por ejemplo la falta de optimización para dispositivos móviles, excesivos tiempos de carga...

3. Proceso de optimización

Comenzaremos el proceso de optimización con una auditoría previa con el fin de conocer su estado de indexación, su estructura, determinar los activos con los que cuenta y encontrar posibles elementos a mejorar.

Posteriormente procederemos a la aplicación de los cambios que nos llevan a adaptar el sitio web a los distintos factores vistos en la Sección 2, y que nos permitirán optimizar el posicionamiento de las páginas del sitio web en los motores de búsqueda.

Dado que como administradores del sitio web solo tenemos control sobre los factores On-Page, nos centraremos en ellos. En cuanto a los factores Off-Page, sobre los que nuestro control es limitado, posteriormente trataremos algún que otro aspecto, pero sin entrar en excesivos detalles. Por último, también hemos visto una serie de factores que afectan de forma negativa a la posición de una página web y que, en nuestro caso, nos limitaremos a tratar de evitar su uso durante este proceso.

3.1. Auditoria previa

En este apartado auditamos el estado previo del sitio web con el fin de conocer su estado de indexación, su estructura, determinar los activos con los que cuenta y encontrar posibles elementos a mejorar.

3.1.1. Estructura y activos del sitio web

El sitio web, como se puede observar en la Figura 9, está estructurado en torno a tres secciones principales, además de la página principales y la página de contacto. Dos de las secciones ('Research' y 'Publications') engloban la práctica totalidad del contenido del sitio web, el principal activo del mismo. Especialmente la sección 'Publications', encargada de enlazar los artículos publicados por el grupo de investigación. En concreto, al inicio de este proyecto el sitio web cuenta con 90 publicaciones, todas ellas en formato PDF.

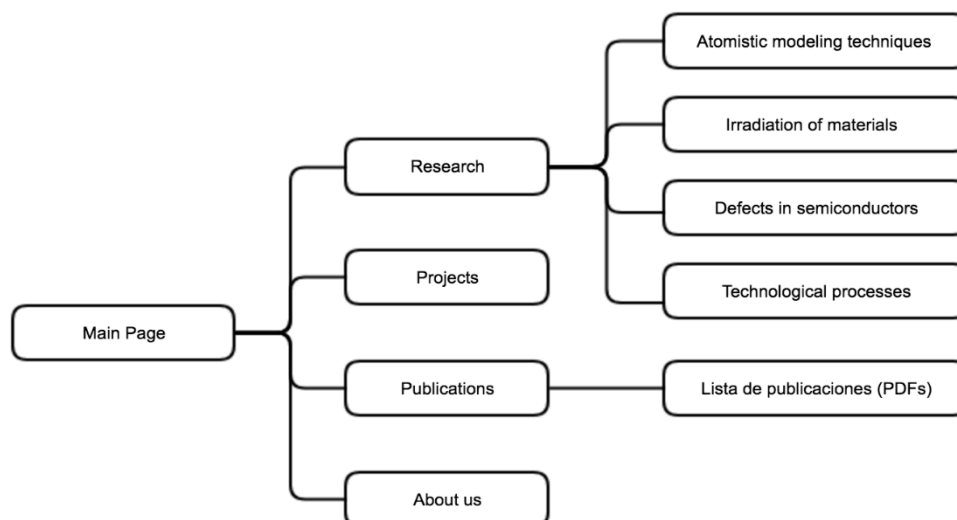


Figura 9. Estructura inicial del sitio web.

Teniendo en cuenta todas y cada una de las páginas web y documentos alojados en el sitio web, este cuenta con un total de 99 páginas web/documentos. La Tabla 5 muestra la forma en que se dividen esas 99 páginas entre las distintas secciones.

Tabla 5. *Páginas del sitio web.*

Sección	Número de páginas
Inicio	1
Research	5(1+4)
Projects	1
Publications	91(1+90 publicaciones PDF)
About us	1
Total	99

Teniendo en cuenta todas y cada una de las páginas web y documentos alojados en el sitio web, este cuenta con un total de 99 páginas web/documentos. La Tabla 5 desglosa esas 99 páginas web en función de

Un sitio web se estructura a través de sus enlaces internos, pues permiten la navegación de los usuarios, ayudan a establecer jerarquía de la información contenida y permiten extender el PageRank a lo largo del sitio web. La estructura de enlaces del sitio web es óptima, mantiene una estructura plana, con un número reducido de clics entre la página principal y cada una de las páginas del sitio web. Tampoco existen contenidos sin enlazar y por tanto los rastreadores de los motores de búsqueda tienen acceso a todo el contenido.

Ahondando en la estructura de la sección 'Research', tenemos que cada una de sus cuatro subsecciones trata a su vez varios temas de interés, que bien podrían contar con su propia página web para hacer un mejor uso del contenido y de las palabras clave que estos temas albergan.

Por último, para mejorar la cohesión y la navegabilidad de la estructura de links cabe considerar el uso elementos como los Breadcrumbs (Migas de pan) y de secciones de "Contenido Relacionado". Especialmente esta última pues en este punto, las publicaciones del grupo de investigación cuentan únicamente con un enlace, en la página principal de la sección 'Publications'. De esta forma distribuiremos mejor el PageRank del sitio web, beneficiando especialmente a las publicaciones.

En este apartado se ha llevado a cabo una simple inspección visual, de la que sin embargo se desprenden aspectos a mejorar, como la reestructuración de la sección 'Research' y la inclusión de elementos que mejoren la estructura interna y su navegabilidad. A medida que se desarrolle el proyecto, especialmente durante el proceso de investigación de palabras clave, se llevará a cabo un análisis más exhaustivo de la estructura, de las palabras clave que le dan forma, y se expondrán con detalle las soluciones aplicadas.

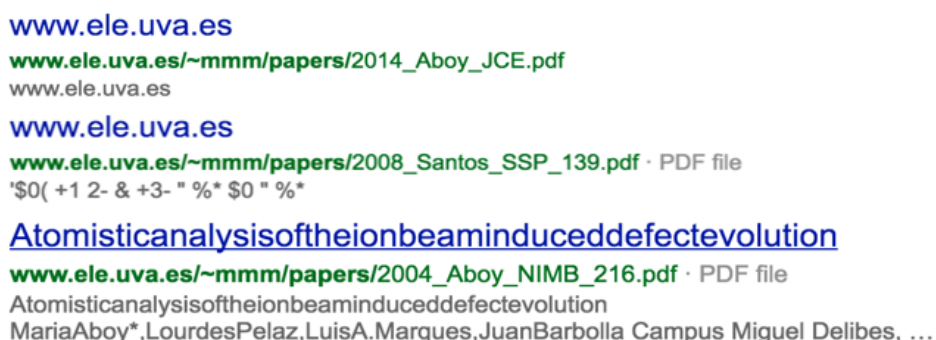
3.1.2. Estado de indexación

Analizando los datos de indexación ofrecidos por la herramienta Google Search Console, en los meses previos al comienzo de este proyecto el sitio web contaba con unos datos de indexación que variaba entre las 60-80 páginas, y teniendo en cuenta que como hemos visto el número total era de 99 páginas web, esto implica que entre un 20-40% de las páginas del sitio web no estaban en el índice de Google y por tanto no eran accesibles desde el buscador.

En el tráfico de búsqueda de Diciembre de 2015, previo al inicio del proyecto, podemos observar aquellas páginas del sitio web que recibieron al menos una impresión, es decir, aparecieron al menos una vez en la página de resultados de una búsqueda. Analizando estos datos podemos ver que todas las páginas web propiamente dichas aparecen en la lista. El problema de indexación por tanto está en los archivos PDF, de los que únicamente recibieron impresiones 51 de ellos, es decir, algo más de la mitad. Muy lejos de lo deseado y en consonancia con los datos revelados anteriormente que situaba en entre un 20-40% las páginas que no estaban indexadas.

En Sección 3.1.1 hemos visto que los archivos PDF se encuentran todos enlazados desde una misma página web, y esta no es una situación ideal y puede ser uno de los problemas del bajo índice de indexación. Un alto número de enlaces en una misma página web puede provocar que el proceso de rastreo de dichas páginas sea más lento. Además, tenemos que, tal y como hemos visto la Sección 1.3.1, un gran número de enlaces implica que el PageRank que una página puede pasar a otras páginas se verá fuertemente fraccionado, diluyendo la importancia y relevancia que dicha página pasa al resto. Hace unos años Google recomendaba que una página debe tener un máximo de 100 enlaces [61]. Actualmente no existe un límite establecido, pero se recomienda hacer un uso coherente.

Por otro lado, nos interesa conocer el grado en el que los motores de búsqueda son capaces de interpretar los archivos PDF, para ello se realizan búsquedas de prueba para observar la apariencia de estos archivos en las páginas de resultados. En la Figura 10 tenemos algunos ejemplos de los problemas que nos hemos encontrado en el aspecto que presentan algunas de las publicaciones en las SERPs, y que pueden ser los causantes de que una parte de ellos no figuren en el índice.



www.ele.uva.es
www.ele.uva.es/~mmm/papers/2014_Aboy_JCE.pdf
www.ele.uva.es

www.ele.uva.es
www.ele.uva.es/~mmm/papers/2008_Santos_SSP_139.pdf · PDF file
'\$0(+1 2- & +3- " %* \$0 " %*

[Atomistic analysis of the ion beam induced defect evolution](#)
www.ele.uva.es/~mmm/papers/2004_Aboy_NIMB_216.pdf · PDF file
Atomistic analysis of the ion beam induced defect evolution
MariaAboy*, LourdesPelaz, LuisA.Marques, JuanBarbolla Campus Miguel Delibes, ...

Figura 10. Ejemplo publicaciones PDF con títulos erróneos en las SERPs.

Estas imágenes dan muestra de que los motores de búsqueda, en concreto Google, se encuentran con problemas para identificar los datos bibliográficos de los archivos PDF. Como se puede ver, en dos de los casos no parece capaz de determinar el título de la publicación y tampoco es capaz de mostrar un fragmento del archivo a modo de descripción. En el caso restante, es capaz de determinar el título y la descripción, pero el formato no es el deseado.

Con ello, se incumplen algunos de los principios del SEO y de la usabilidad, tanto para los usuarios como para los motores de búsqueda. Si los archivos no cuentan con un título y una descripción que sean descriptivas del contenido, difícilmente el motor de búsqueda indexará dicho archivo para las keywords que le correspondan. Y si lo hiciera difícilmente un usuario visitará una página de la que desconoce su título y su descripción. Por tanto, el problema de los PDF afecta también a la experiencia de usuario.

La solución propuesta para facilitar la interpretación de los datos bibliográficos por parte de los motores de búsqueda pasa, de momento, por analizar los metadatos de cada uno de los archivos PDF, y añadir/modificar los datos que sean pertinentes. De esta forma, una vez que los motores de búsqueda interpreten correctamente dichos datos bibliográficos mejorará el formato que presentan los archivos PDF en las SERPs y su indexación.

3.1.3. Análisis de competidores

Nuestro objetivo principal en este proyecto es mejorar el posicionamiento de las páginas del sitio web del Multiscale Materials Modeling Group, en especial de su página principal.

En el caso de la página principal, nuestros principales competidores serán aquellos sitios web que posicionan contenido para términos como *'multiscale materials modeling'*, *'materials modeling'*, *'multiscale materials'*... Analizando las SERPs para esos términos tenemos que se trata de sitios web de otros grupos de investigación, sitios web dedicados a las conferencias MMM realizadas anualmente, entradas de Wikipedia, libros, artículos...

En principio los competidores más duros son Wikipedia, debido a que se trata de un dominio de gran autoridad y prestigio, y los sitios web dedicados a las conferencias anuales. Este último se debe a que, aunque se trata de sitios web de carácter estacional, es decir, su utilización está limitada a un periodo de tiempo en torno a las fechas de la conferencia, reciben una gran atención en forma de enlaces y menciones en redes sociales. Esto lleva a que los sitios web de conferencias MMM correspondientes a los años 2010 y 2012, aún continúen obteniendo grandes resultados un lustro después.

Por otro lado, como hemos visto a la hora de revisar la estructura y los activos del sitio web, su principal activo son las publicaciones llevadas a cabo por el Multiscale Materials Modeling Group. En concreto suponen alrededor del 90% del contenido que debemos posicionar. Por tanto, nuestros principales competidores son las páginas de las revistas científicas que han publicado dichos artículos.

Debemos ser conscientes de la dificultad que esto supone a la hora de posicionar las publicaciones, pues estamos tratando de posicionar exactamente el mismo contenido. Además, hay que tener en cuenta que, tal y como funciona el sistema de citaciones y referencias de los artículos académicos, cualquier citación que se realice a un artículo se realiza a través de su DOI (Document Object Identifier) que enlaza con la página de la revista científica. Es decir, las páginas de las publicaciones reciben todos y cada uno de los enlaces, lo que les da una posición privilegiada frente al resto de competidores.

Analizando las páginas de las distintas publicaciones y como estas aparecen en las SERPs, podemos observar que cada publicación cuenta con dos versiones de la misma. Por una parte, cuenta con una página HTML que contiene la información bibliográfica del artículo (Título, autores, nombre de la publicación, fecha...) así como una pequeña descripción (abstract). Además, incluye otra serie de funcionalidades con las que no cuentan los archivos PDF, como compartir el artículo a través de redes sociales, acceder a artículos relacionados... Finalmente esta página HTML nos proporciona el acceso, mediante pago o suscripción, al archivo PDF que contiene el artículo completo. Es decir, el mismo archivo que proporcionamos actualmente desde nuestro sitio web.

Esto nos proporciona una posible solución a los problemas que nos generan los archivos PDF a la hora del SEO, y que puede ser los causantes del irregular índice de indexación de los archivos PDF. Cuando revisamos el estado de indexación ya vimos los problemas que nos causaba el hecho de tener poco control sobre el aspecto que presentan los archivos PDFs en las SERPs. En el apartado reservado a la auditoría del aspecto visual de la página profundizaremos en la aplicación de esta solución.

3.1.4. Aspectos visuales que mejorar

Se ha llevado una revisión visual de las páginas del sitio web para determinar si existen elementos, como pueden ser títulos, descripciones, enlaces, imágenes..., que necesiten o puedan ser mejorados.

Se han detectado errores a la hora de utilizar las etiquetas de título, las descripciones, en las etiquetas de encabezado, anchor text, ... La importancia de alguno de estos errores es relativa desde el punto de vista del SEO, sin embargo, desde el punto de vista de la usabilidad y de la estética cobran mayor importancia. En la Figura 11 podemos ver un ejemplo que ilustra perfectamente el problema.

[Multiscale Materials Modeling Group @ UVa website :: Research :: Irr...](http://www.ele.uva.es/~mmm/research.irradiation.html)
www.ele.uva.es/~mmm/research.irradiation.html
Main results on irradiation of materials of the Multiscale Materials Modeling research group of the University of Valladolid.

Figura 11. Ejemplo etiqueta título incorrecta.

Se trata del formato de las etiquetas de título que originalmente presenta el sitio web. Como podemos observar los motores de búsqueda truncan el título por ser excesivamente largo y el tema sobre el que trata la página queda oculto.

También se han detectado errores a la hora de emplear los 'anchor text' de los enlaces. En la Figura 12 podemos ver que se emplea el mismo 'anchor text' para todos los archivos PDF.

Insights on the atomistic origin of X and W photoluminescence lines in c-Si from ab initio simulations
I. Santos, M. Aboy, P. López, L. A. Marqués, and L. Pelaz
[Journal of Physics D: Applied Physics 49, 075109 \(2016\), \[PDF\]](#)

Molecular dynamics simulation of the early stages of self-interstitial clustering in silicon
L. A. Marqués, I. Santos, L. Pelaz, P. López, and M. Aboy
[Materials Science in Semiconductor Processing 42, 235 \(2016\), \[PDF\]](#)

Atomistic modeling of ion implantation technologies in silicon
L. A. Marqués, I. Santos, L. Pelaz, P. López, and M. Aboy
[Nuclear Instruments and Methods in Physics Research B 352, 148 \(2015\), \[PDF\]](#)

Figura 12. Utilización incorrecta del anchor text.

Es conveniente que el ‘anchor text’ del enlace aporte información acerca del contenido al que dirige tanto a los usuarios como a los motores de búsqueda. Por lo tanto, se procederá a la corrección de este pequeño error.

El problema de los archivos PDF descrito en las Sección 3.1.2 es únicamente uno de los inconvenientes que presentan este tipo de archivos dentro del SEO. Además de reducir nuestro control sobre la información que se muestra en las SERPs tenemos que reduce al mínimo la interacción de usuario con el resto del sitio web. Desde el punto de vista SEO nos conviene que el usuario pase el mayor tiempo posible en nuestro sitio web, y esto se lleva a cabo proporcionando al usuario la posibilidad de visitar otras secciones del sitio web, compartiendo contenido vía redes sociales... algo imposible de llevar a cabo desde un PDF.

En el análisis de competidores hemos visto que las revistas y asociaciones científicas en las que se publican los artículos del grupo de investigación cuentan con una página HTML para cada una de las publicaciones, en la que se incluyen los datos bibliográficos fundamentales, así como una breve descripción (abstract), y desde la que se da acceso a la versión completa del artículo.

Por tanto, este método se ajusta a nuestras necesidades, pues nos permite tener un mayor control tanto de la información mostrada a los usuarios como de los aspectos estéticos. Además, permite a los usuarios interactuar con el resto del sitio web de forma normal, al contrario de lo que ocurre con los PDFs. Por último, nos permite añadir funcionalidades como la posibilidad de compartir el contenido a través de redes sociales, elementos de interés como las secciones de “Publicaciones relacionadas”, que permiten a los usuarios descubrir otras publicaciones del sitio web y de esta forma pasar un mayor tiempo visitando el sitio web.

3.1.5. Posicionamiento previo

Por último, analizaremos la posición que ocupan en el ranking las principales secciones del sitio web, para sus respectivas palabras clave. Prestaremos especial atención a la sección ‘Research’, sus subsecciones y los distintos temas tratados en cada una de ellas, pues suponen uno de los principales activos del sitio web junto con las publicaciones, en las que no entraremos en detalle.

El fin de este análisis es determinar la visibilidad de dicho contenido. Y en el caso de que este no sea visible tomar las medidas necesarias. En este caso se ha llevado a cabo una búsqueda manual en Google y Bing, a diferencia del proceso de seguimiento del proyecto para el que se empleará un

software que automatice la tarea. La Tabla A.1. del Anexo nos muestra el posicionamiento previo de las principales palabras clave del sitio web, y de sus correspondientes páginas.

Como es natural, dado que emplean unas palabras clave demasiado comunes, las páginas principales de secciones como 'Research', 'About us', 'Publications' y 'Projects' no aparecen en los resultados de búsqueda. Para ello los usuarios tendrán que realizar búsquedas mucho más específicas, como por ejemplo empleando el nombre de la universidad o del grupo de investigación en su consulta.

En cuanto a la página principal del sitio web, aparece en posiciones irrelevantes para la palabra clave objetivo (*Multiscale materials modeling*), tanto en Google (32) como en Bing (53). Uno de los objetivos principales de este proyecto es mejorar la posición de la página principal para dicha palabra. Ahora, conocemos el punto de partida.

Por último, y como se puede ver en la Tabla A.1., tanto Google como Bing indexan correctamente las páginas correspondientes a las subsecciones de la sección 'Research' (excepto la subsección *Technological processes*) para sus respectivas palabras clave. En cambio, presentan serias dificultades a la hora de indexar dichas páginas para los términos correspondientes a los diferentes temas sobre los que tratan. El hecho de que cada una de las subsecciones albergue varios temas diferentes puede hacer que, a ojos de los motores de búsqueda, la relevancia del contenido de cada uno de los temas se vea diluida, y por tanto no aparezca en los resultados para sus palabras clave.

Con estos datos, la reestructuración de la sección 'Research', mencionada anteriormente en esta auditoría, resulta indispensable para mejorar la visibilidad del contenido y hacer un uso más eficiente de las palabras clave con las que contamos. Además, dado que las páginas de la sección 'Research', aun siendo indexadas, no ocupan lugares de importancia, el riesgo de perder posiciones para dichas palabras clave en una reestructuración, frente a la posibilidad de mejorar notablemente la posición para otras palabras clave, resulta asumible.

A modo de resumen, la auditoría previa nos deja una serie de cambios principales que debemos realizar.

- Reestructuración de la sección 'Research': Una página web por cada uno de los temas tratados en las subsecciones.
- Reestructuración de la sección 'Publications': Una página web por cada publicación.
- Uso de elementos como Breadcrumbs o secciones de 'Contenidos relacionados' con el fin de mejorar la cohesión entre los contenidos del sitio web y redistribuir el PageRank del sitio web, beneficiando especialmente a las publicaciones que obtendrán mayor relevancia.
- Revisión de los metadatos de los archivos PDF correspondientes con las publicaciones.

3.2. SEO On-Page

El SEO On-Page agrupa a aquellos factores que están bajo el control del webmaster del sitio web. Por tanto, se trata de factores sobre los que podemos influir directamente.

En este apartado llevaremos a cabo las modificaciones que resulten pertinentes para adaptar el sitio web a los requerimientos impuestos por los distintos factores. El objetivo de estas modificaciones no es otro que optimizar el posicionamiento de las páginas del sitio web en los motores de búsqueda.

Por un motivo de coherencia mantendremos el mismo formato que vimos cuando descubrimos los factores SEO, en la sección 2.1. Por tanto, dividiremos las modificaciones en función de la tipología del factor al que afecta: contenido, arquitectura y aspectos técnicos, y código HTML.

3.2.1. Contenido

El contenido es el activo más importante de cualquier sitio web, y sienta las bases del éxito de una campaña SEO. Sin un contenido de calidad resultará difícil que el resto de cambios involucrados en el proceso obtengan los resultados esperados. En este apartado describiremos los cambios aplicados relativos a los factores que afectan al contenido.

3.2.1.1. Contenido de calidad, fresco y único

El contenido es el activo más importante de cualquier sitio web, y sienta las bases del éxito de un proyecto SEO. Sin un contenido de calidad, fresco y único, resultará difícil que el resto de cambios involucrados en el proceso obtengan los resultados esperados. Además, un contenido con estas características resultará más relevante tanto para los motores de búsqueda como para los usuarios, haciendo que este atraiga más visitas, más enlaces, ... y redunde en una mejora de la posición en las páginas de resultados.

La calidad del contenido del sitio web objeto de este proyecto está fuera de toda duda. En su mayoría se trata de artículos académicos, escritos por los miembros del grupo de investigación y publicados en prestigiosas revistas científicas como Elsevier, American Physical Society, American Institute of Physics...

Por otro lado, tenemos la frescura. Dada la temática del sitio web, la importancia de la frescura del contenido es relativa. Existen artículos académicos de hace 20 años cuya vigencia hoy en día es absoluta y otros más actuales, en cambio, se encuentran desfasados. Por tanto, no se trata tanto de un tema de frescura como de la vigencia de los artículos. Aun así, existirá un periodo de tiempo durante el cual los nuevos artículos verán incrementado su posición debido a su frescura.

En cuanto a la unicidad del contenido, tenemos un problema. Las páginas de las revistas científicas, que publican los artículos del grupo de investigación, ofrecen exactamente el mismo contenido y, de hecho, son nuestro principal competidor. Por lo tanto, desde el punto de vista de los motores de búsqueda, el contenido no es único.

Si bien, en este caso no se trata de una cuestión de plagio, robo de contenido..., susceptibles de ser penalizadas, pues son los propios autores de los artículos los que los publican en el sitio web. Se trata más bien de ofrecer una fuente alternativa para dichos artículos, y este parece ser también el punto de vista de los motores de búsqueda, pues han considerado que los artículos incumplan sus normativas.

3.2.1.2. Investigación de palabras clave

Antes de comenzar con el proceso de investigación propiamente dicho expondremos algunos de los aspectos que afectan al proceso y que debemos considerar. Además, describiremos las técnicas y herramientas empleadas durante el proceso.

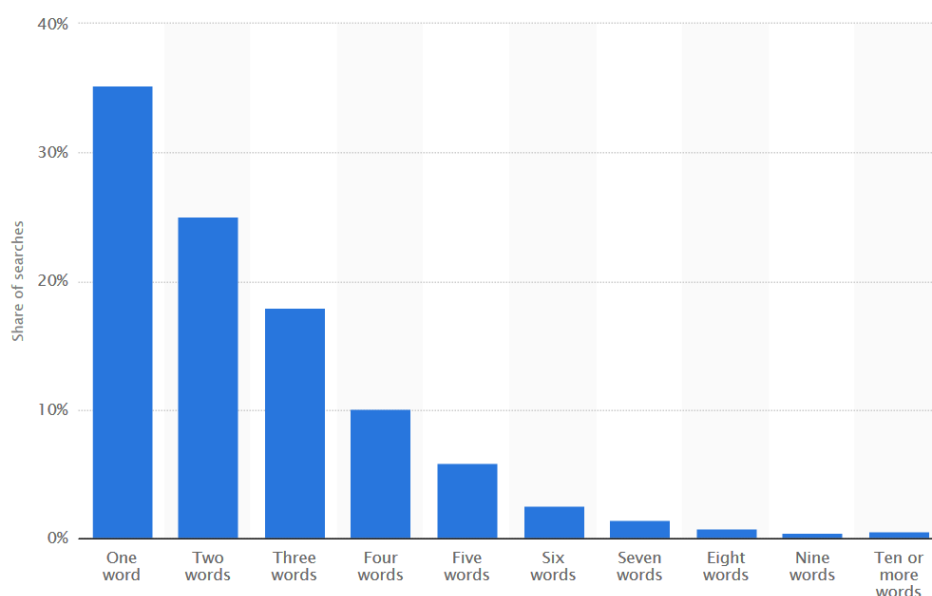
3.2.1.2.1. Aspectos previos

- ¿Que es una Keyword?

Las keywords o palabras clave son uno de los pilares del SEO. Son las palabras, o combinaciones de las mismas, que emplean los usuarios en sus consultas. Resulta de vital importancia conocer cuáles son los términos utilizados por los usuarios a la hora de buscar el contenido de un sitio web, pues de esta forma podremos conocer si el contenido del mismo está enfocado a dichos términos y, en el caso de no ser así, realizar las modificaciones pertinentes. Además, una vez conocidas las palabras clave podremos enfocar hacia dichas palabras la generación de nuevos contenidos.

De poco sirve tener un sitio web con contenido de calidad, si los términos para los que posiciona no son los correctos dentro de ese mercado o, aun siendo correctos, no son los que los usuarios están empleando en sus consultas y por tanto estamos perdiendo potenciales usuarios por no emplear las palabras clave adecuadas.

Antes de comenzar a relatar el procedimiento llevado a cabo para la selección de las palabras clave empleadas en el proyecto explicaremos brevemente la situación actual de la investigación de palabras clave. La Figura 13 muestra el porcentaje de búsquedas en función del número de palabras empleadas. Los datos corresponden a Febrero de 2017 para el mercado de EE. UU. [62]. Como podemos observar predominan las búsquedas cortas (1 o 2 palabras), que aglutinan cerca del 60% de las búsquedas.



© Statista 2017

Figura 13. Porcentaje de búsqueda en función del número de palabras clave. Statista 2017 [62]

Sin embargo, conviene hacer un análisis de la demanda de dichos términos, pues si bien es cierto que estos engloban cerca del 60% de las búsquedas, esto es gracias a que la mayoría de las palabras clave más populares están formadas por una o dos palabras.

La Figura 14 muestra la situación en términos de demanda de keywords. Como se puede ver, las palabras clave más populares concentran cerca del 30% del tráfico global. Se trata de palabras clave cortas (1 o 2 términos). La situación ideal es clasificar para alguna de estas palabras clave, pues reciben decenas de miles de búsquedas mensuales. El inconveniente es que la competencia y el coste tanto monetario (en el caso de una campaña PPC) como de recursos por posicionar una página para estas palabras clave será muy alto.

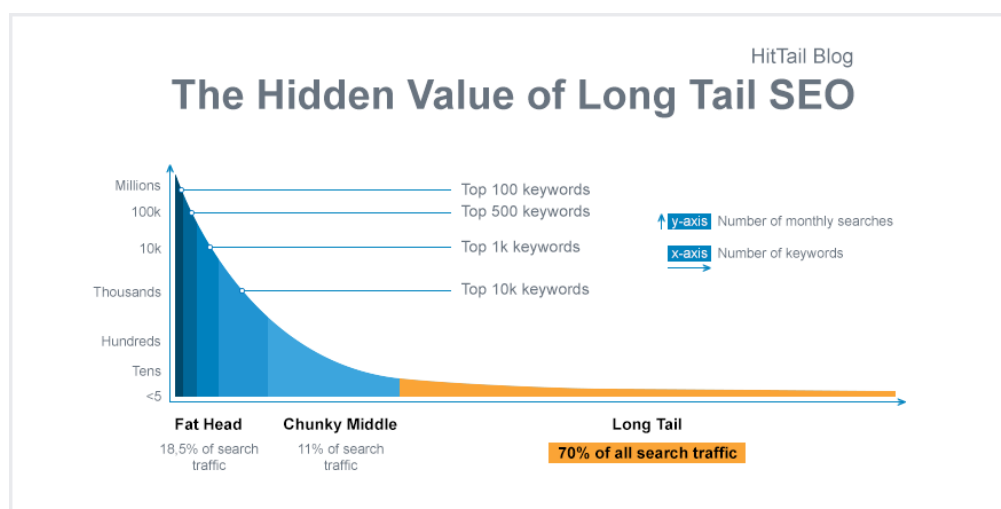


Figura 14. Demanda en función del número de palabras clave. 'Long tail'.

Por otro lado, tenemos que el 70% restante, denominado 'long tail'. El 'long tail' contiene cientos de millones de búsquedas únicas, y aunque reciban un número muy reducido de búsquedas mensuales, cuando se evalúan en conjunto suponen una amplia mayoría de la demanda de información que reciben los motores de búsqueda. La competencia y el coste por posicionar las keywords que forman parte del 'long tail' es mucho más baja. Y la probabilidad de conversión (clic en nuestra web) aumenta pues el usuario está realizando búsquedas mucho más específicas.

- Hummingbird

Como vimos en la sección 1.3, en la que repasamos las grandes actualizaciones del algoritmo de Google, en Septiembre de 2013, y coincidiendo con su 15º aniversario, Google anunció una de sus mayores actualizaciones, denominada Hummingbird. Esta actualización supone un cambio en la forma en la que Google interpreta las búsquedas y en cómo las búsquedas se relacionan entre sí.

El cambio viene para adaptar el motor de búsqueda a una nueva situación, en la que las búsquedas desde dispositivos móviles comienzan a ocupar una parte significativa del mercado. En este tipo de búsquedas, debido en parte a las condiciones propias de los dispositivos móviles, los usuarios tienden a realizar consultas con un formato más críptico, alejado del formato tradicional, el del escritorio. Además, en la actualidad están cobrando importancia las búsquedas mediante la voz, lo que les confiere a las consultas un carácter y un formato conversacional (Ej. Encuentra el hospital más cercano).

Todos estos cambios están orientados a determinar cuál es la verdadera intención del usuario cuando realiza una consulta, en contraposición al método anterior en el que se ofrecen los resultados en base al contenido exacto de dicha consulta.

Este enfoque no solo afecta a la forma en la que Google interpreta las consultas realizadas por los usuarios, también afecta a la forma en la que interpreta las páginas web. Desde este enfoque, Google trabaja para comprender el significado y la temática detrás del contenido de la página, en lugar de tener en cuenta únicamente las palabras clave incluidas en el contenido de las mismas. De esta forma una página no clasifica en función de su optimización para una palabra clave determinada, clasifica en función del contenido, y de lo que este se ajuste a la intención de la consulta realizada por el usuario [63].

- **Concurrencia**

En este escenario cobra importancia el concepto de análisis de concurrencia (co-ocurrencia), es decir, el análisis del contenido de la página con el fin de determinar qué palabras y frases aparecen con mayor frecuencia, con qué frecuencia aparecen, y qué relación existe entre ellas. Utilicemos un ejemplo para observar de forma más clara el concepto de concurrencia, su aplicación al análisis de páginas web y cómo es empleado por Google.

Tenemos un comercio online dedicado a la venta de libros. Las palabras *libro*, *ciencia-ficción* y *Asimov* coexisten en la página con palabras propias de un comercio online como son *carro*, *envío*, *descuento*, *venta*... La frase '*comprar libros ciencia-ficción online*' puede que no aparezca en toda la página web y, sin embargo, ante la concurrencia de los términos propios de los productos en venta (libros) y de los términos propios de un comercio online, el motor de búsqueda interpreta que el propósito de la página web es la venta de libros, y potencialmente la página puede que llegue a clasificar para dichos términos. Aunque esto también dependerá de otros muchos factores.

El objetivo del motor de búsqueda no es otro que entender la intención del usuario cuando este realiza la búsqueda y devolver el conjunto de resultados que satisfaga dicha intención, independientemente de si los términos exactos de la consulta aparecen en el contenido de las páginas [63].

- **Encriptación de las consultas (Not provided/Not set)**

En la sección 1.3, en la que repasamos las actualizaciones más importantes del algoritmo de Google, hablamos de la actualización Query Encryption. En ese momento comentamos que dicha actualización supuso un duro golpe para el SEO, en especial a la investigación de palabras clave. A continuación, veremos los motivos.

Hasta el momento de la actualización, el webmaster de una web a la que se accedía a través de los resultados de una búsqueda orgánica en <http://www.google.com> (sin el protocolo de encriptación SSL), tenía conocimiento de que dicha visita procedía de google.com y conocía la consulta que había provocado esa visita. Sin embargo, tras la actualización, tenemos que para aquellas visitas que proceden de búsquedas orgánicas en las que se ha empleado el protocolo SSL, el webmaster solo tiene conocimiento de que dichas visitas provienen de google.com. Las consultas que provocaron dichas visitas aparecerán ahora en las estadísticas bajo un enigmático "not provided".

Este hecho dificulta notablemente muchos procesos que forman parte del SEO: la investigación de palabras clave, el seguimiento de las visitas en función de la consulta, la posibilidad de conocer las consultas que los usuarios emplean para llegar a tu sitio web y en caso de ser necesario adaptar el contenido a dichas consultas...

Además, tenemos otro inconveniente con la etiqueta “not set”. Según se indica en la herramienta Google Analytics el objetivo de esta es “garantizar la privacidad de los usuarios”, por ello “las consultas que se realizan muy pocas veces, o que contienen información personal o confidencial, se agrupan en la fila (not set)”. Precisamente nos encontramos en la situación en que nuestras consultas objetivo reciben un número de visitas muy bajo, como veremos al final del proceso de investigación de palabras clave.

En la Figura 15 se puede ver el aspecto con el que aparece la información en Google Analytics tras la incorporación de “not provided” y “not set”.

Consulta	Impresiones ? ↓	Clics ?	Posición media ?	CTR ?
	2.012 % del total: 100,00 % (2.012)	61 % del total: 100,00 % (61)	27 % del total: 100,00 % (27)	3,03 % Media de la vista: 3,03 % (0,00 %)
1. (no definido)	1.594 (79,22 %)	59 (96,72 %)	25 (89,20 %)	3,70 %
2. multiscale materials modeling	48 (2,39 %)	1 (1,64 %)	27 (97,63 %)	2,08 %
3. multiscale materials	39 (1,94 %)	0 (0,00 %)	34 (126,28 %)	0,00 %
4. atomistic modeling	34 (1,69 %)	0 (0,00 %)	42 (155,33 %)	0,00 %
5. defects in semiconductors	32 (1,59 %)	0 (0,00 %)	44 (163,74 %)	0,00 %
6. materials modeling	29 (1,44 %)	0 (0,00 %)	61 (226,17 %)	0,00 %

Figura 15. Ejemplo “not provided”, “not set”.

- Canibalización de palabras clave

La canibalización de palabras clave se produce cuando varias páginas de un mismo sitio web comparten una misma temática, hacen uso de las mismas palabras clave y por tanto compiten entre ellas a la hora de conseguir una posición prominente en el ranking para dichas palabras clave.

Generalmente los buscadores tratan de incluir cierta diversidad en sus páginas de resultados, tanto de contenidos como de fuentes. Por ello, cuando un sitio web cuenta con múltiples páginas orientadas hacia una misma keyword, el motor de búsqueda se ve forzado a seleccionar aquella que considere más relevante, y en muchos casos, rechazando el resto, de forma que estas no se muestran en las páginas de resultados.

3.2.1.2.2. Técnicas

Las técnicas que describiremos a continuación nos permiten obtener un primer conjunto de palabras clave que podremos completar con el uso de diversas herramientas que veremos más adelante.

- Papel y lápiz

En primer lugar, emplearemos una de las técnicas tradicionales a la hora de obtener nuevas ideas o bien conocer las ya existentes, el brainstorming o lluvia de ideas. Con la ayuda de los participantes en el negocio empezaremos por generar una lista de términos que son relevantes dentro del sector de la empresa, añadiremos también los términos que se considera que el sitio web incluye actualmente. Esta lista debería tener una extensión que va desde varias decenas a varios centenares de términos, dependiendo del tamaño del negocio y del sitio web.

En el caso de una empresa es importante que el puesto y la ocupación de los participantes en el proceso sea lo más variado posible, con el fin de abarcar todos los ámbitos en los que opera la empresa y obtener con la mayor precisión la terminología utilizada en el negocio. En este aspecto es importante contar con la ayuda del departamento de ventas, marketing, servicio técnico..., es decir, con aquellas personas que mantengan un contacto directo con el cliente.

Este proceso nos permitirá conocer el concepto que se tiene desde dentro de la propia empresa y el lenguaje empleado por la misma para ofertar sus productos/servicios. Además, comenzaremos a descubrir aquellos términos que emplean los clientes a la hora de referirse a los productos/servicios ofrecidos por la empresa, y que puede que aún no estuvieran en la lista.

Los pasos a seguir en el proceso de brainstorming son los siguientes [\[64\]](#):

- Realizar una lista inicial de palabras clave (longitud 1-3) que describan los productos/servicios ofertados.
- Encontrar sinónimos que pudieran ser utilizados por los clientes a la hora de referirse a dichos productos/servicios.
- Realizar una taxonomía de las áreas de intereses dentro del sector. Observar la jerarquía de los sitios web de las empresas líderes en el sector.
- Incluir nuevos términos generales en los que los productos/servicios ofertados por la empresa pudieran ser incluidos.
- Analizar el sitio web actual y extraer las palabras clave empleadas.
- Listar todos los nombres de marca de la empresa.
- Listar todos los nombres de los productos. Si el número fuera muy elevados considerar categorías y subcategorías e incluirlas en la lista.
- Imaginar que somos consumidores potenciales y preguntarnos qué consultas realizaríamos en un motor de búsqueda si deseáramos encontrar productos/servicios similares a los ofertados por la empresa.
- Complementar el paso anterior realizando la misma pregunta a personas ajenas a la empresa.
- Emplear herramientas para analizar cuáles son los términos que emplean los clientes que visitan el sitio web en la actualidad.

- **Análisis de competidores**

El objetivo del análisis de competidores es determinar qué palabras clave incluyen los sitios web de dichos competidores y cómo estas encajan en la estructura de los mismos, determinando contenidos y secciones. Para ello analizaremos aquellos sitios web que supongan un referente en el campo en que esté situado el sitio web objeto de nuestro trabajo.

Este análisis ha de servir para complementar los resultados obtenidos en el proceso descrito anteriormente, permitiéndonos dar cuenta de aquellos términos que en el caso de nuestros competidores son relevantes para su posicionamiento y, en el caso de que estos no fueran incluidos con anterioridad en nuestra lista de palabras clave, incluirlos.

En nuestro caso no existen competidores propiamente dichos, pues se trata de un grupo de investigación y el objetivo de este proyecto es mejorar el acceso a los contenidos del sitio web de grupo desde los motores de búsqueda.

Aun así, analizaremos los competidores más probables teniendo en cuenta el contenido del sitio web. Como vimos en la auditoría inicial, el principal activo del sitio web son las publicaciones realizadas por el grupo de investigación. Todas ellas cuentan con un enlace al a la página web de la revista científica donde fue publicado, y donde por un módico precio puede ser adquirido. Por tanto, trataremos de analizar el formato y las palabras clave empleadas por las revistas científicas en sus páginas web con el complementar la lista de términos [\[65\]](#).

3.2.1.2.3. Herramientas

Podemos obtener datos de utilidad para la investigación de palabras clave de diversas fuentes, tanto de funcionalidades incluidas en los propios buscadores como en herramientas independientes. En este último caso la oferta es abundante, existiendo alternativas gratuitas como de pago.

- Funcionalidades de los propios buscadores.

En primer lugar, los propios motores de búsqueda cuentan con funcionalidades que nos pueden ayudar a obtener nuevos términos a considerar.

- *Términos relacionados.* La mayoría de los motores de búsqueda ofrecen términos relacionados con las keywords de la búsqueda realizada. Esta funcionalidad puede proporcionar nuevos términos que no surgieron con el brainstorming ni con el análisis de competidores.
- *Uso del operador *.* Empleado junto con una palabra clave nos permite conocer qué términos preceden o siguen normalmente a la palabra clave objeto de nuestra búsqueda.
- *Función Autocompletar.*

- Google AdWords Keyword Planner

Esta herramienta está diseñada específicamente para su uso en la investigación de palabras clave [\[66\]](#). En principio está destinada a los usuarios de Google AdWords, la plataforma de publicidad de pago por clic (PPC) de Google, pero puede utilizarse para obtener información para las búsquedas orgánicas. La herramienta proporciona términos relacionados, estimaciones sobre los volúmenes de búsqueda, tendencias de búsqueda, y estimaciones de coste en una campaña PPC para cualquier palabra clave o URL que se introduzca en la herramienta. El uso de la herramienta no conlleva coste alguno, pero está sujeto a la necesidad de contar con una cuenta en la plataforma Google Adwords.

Con el fin de encontrar nuevas palabras clave y obtener datos del volumen de búsqueda, la herramienta ofrece tres modos de utilización:

- Buscar palabras clave nuevas mediante una palabra clave/frase, sitio web o categoría.
- Obtener datos y tendencias del volumen de búsqueda (a partir de una lista de palabras)
- Combinar listas de palabras clave para obtener palabras clave nuevas.

En nuestro caso únicamente nos interesa el funcionamiento del primero de estos modos, ya que a partir de una palabra clave/frase, un sitio web o una categoría, nos ofrece nuevas palabras clave de interés.

El funcionamiento es muy simple, basta con introducir en la herramienta una palabra clave, un sitio web o una categoría sobre las que deseamos buscar nuevas palabras clave. En la Figura 16 podemos visualizar el resultado de una búsqueda.

Su producto o servicio

multiscale materials modeling

Obtener ideas Modificar búsqueda

Ideas para el grupo de anuncios Ideas para palabras clave Columnas Descargar Añadir todas (296)

Términos de búsqueda	Promedio de búsquedas mensuales	Competencia	Puja sugerida	Porcentaje de impresiones del anuncio	Añadir al plan
multiscale materials modeling	40	Baja	–	–	CUENTA

Mostrar filas: 30 1 - 1 de 1 palabras clave < > >> <<

Palabra clave (por relevancia)	Promedio de búsquedas mensuales	Competencia	Puja sugerida	Porcentaje de impresiones del anuncio	Añadir al plan
multiscale	2.400	Baja	–	–	»
multiscale modeling and simulation	170	Baja	–	–	»
modelling materials	320	Baja	0,16 €	–	»
engineering scale	1.300	Media	0,52 €	–	»

Figura 16. Google AdWords Keyword Planner.

Como podemos ver se ofrecen un conjunto de palabras clave junto con su promedio de búsquedas mensuales, la competencia y una cifra orientativa del precio que podría tener una campaña de pago por clic para dicha palabra clave.

Además, contamos con una serie de filtros que nos permiten adaptar los resultados de la búsqueda en función de su ubicación, idioma, redes de búsqueda empleadas, periodo de tiempo y la opción de incluir/excluir palabras clave concretas.

Como podemos ver el Planificador de palabras clave es realmente útil pues nos proporciona una gran cantidad de datos que nos permite encontrar nuevas palabras clave, estimaciones sobre la popularidad de dichas palabras claves, volumen de búsqueda, ... con la ventaja de que los datos los proporciona la propia Google y son obtenidos de las búsquedas realizadas en el propio buscador.

- Bing Keyword Research

Incluida en la plataforma Bing Webmasters Tools, la herramienta Bing Keyword Research proporciona sugerencias sobre palabras clave basadas en los términos introducidos en la herramienta. Así mismo nos proporciona el volumen de búsquedas para dichos términos y una estimación del coste en una campaña de pago por clic en la plataforma Bing Ads [67].

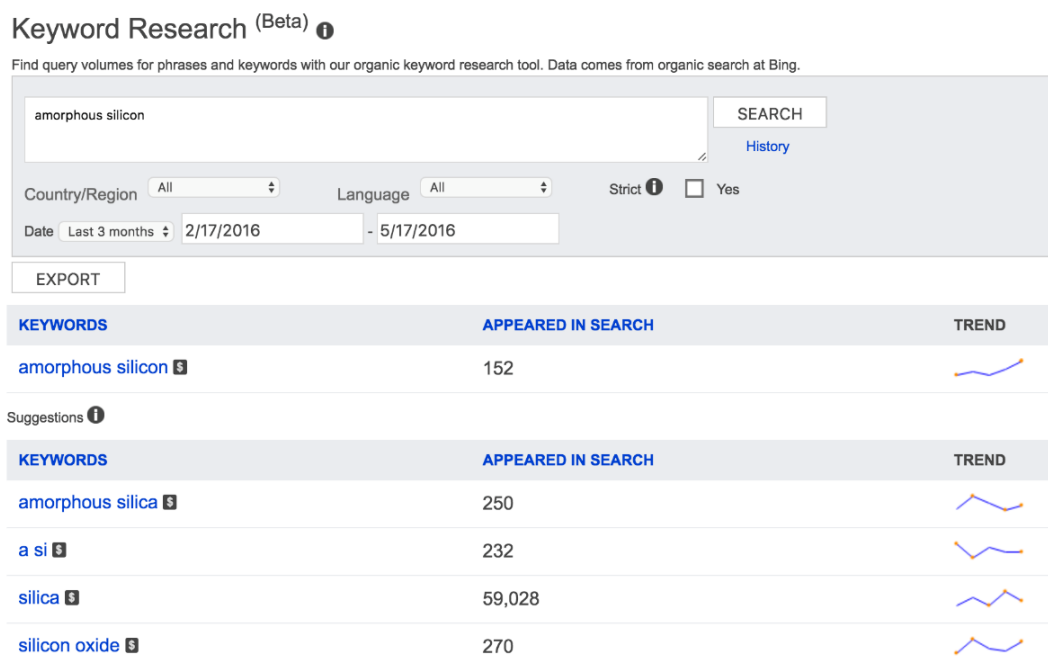


Figura 17. Bing Keyword Research.

Como se puede ver en la Figura 17 la herramienta nos permite filtrar los resultados en función de un rango de fechas, la ubicación y el idioma.

Los datos proporcionados por la herramienta se obtienen de la propia base de datos de Bing, por lo tanto, la información es sumamente valiosa pues nos permite conocer de primera mano qué palabras clave tienen mayor relevancia para el público, que palabras tienen mayor competencia a través de las estimaciones de coste por clic, determinar tendencias estacionales de dichas palabras... y de esta forma podremos enfocar nuestros esfuerzos mucho mejor. El uso de la herramienta no conlleva coste alguno, pero está sujeto a la necesidad de contar con un sitio web registrado en la plataforma Bing Webmasters Tools.

- Google Trends

Como su propio nombre indica Google Trends nos permite visualizar las tendencias seguidas por la popularidad de las palabras clave a lo largo del tiempo y por su distribución geográfica. La herramienta nos permite realizar comparaciones entre palabras clave pudiendo de esta forma comparar su popularidad y determinar tendencias estacionales a lo largo del tiempo [68].

Google Trends no aporta cifras absolutas del volumen de búsqueda de las palabras clave, como sí lo hacían las herramientas descritas anteriormente. El formato de sus resultados es mucho más gráfico, lo que nos permite con un simple vistazo ver las tendencias seguidas a lo largo del tiempo, tal y como podemos ver en la Figura 18.

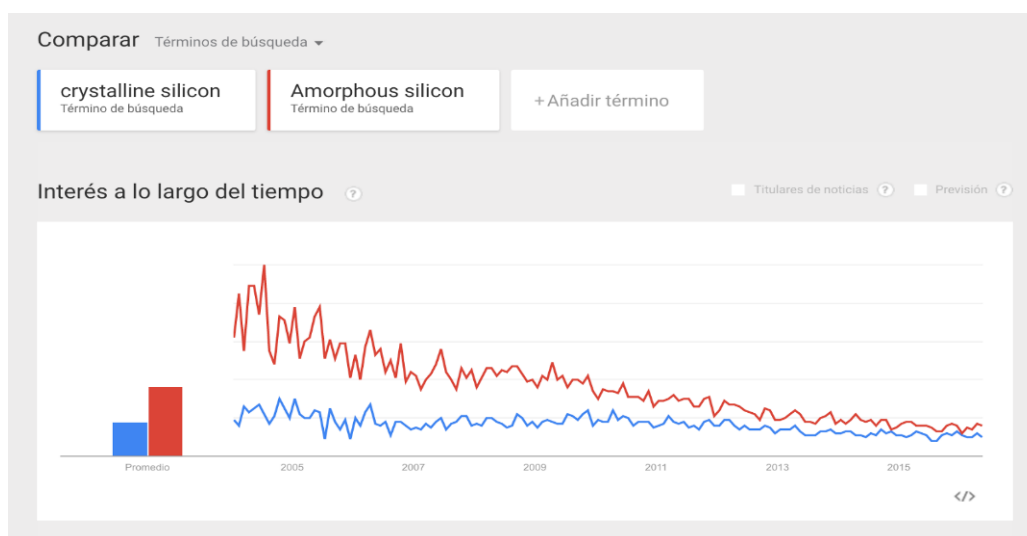


Figura 18. Google Trends.

Además, nos permite hacer un desglose geográfico para determinar qué ciudades, países y en qué lenguajes se han realizado la mayoría de las consultas para una palabra clave. Por último, proporciona un conjunto de “Búsquedas relacionadas” con los términos empleados por el usuario.

Google Trends es una herramienta gratuita, y sus datos proceden de las búsquedas realizadas en el propio motor de búsqueda de Google.

- **Keyword Density Analyzer**

Nos permite analizar un sitio web existente, un texto... y desglosar su contenido en función de la utilización de las palabras en el mismo. De esta forma podemos conocer qué palabras se repiten con mayor frecuencia, es decir tienen mayor densidad, y por tanto determinar qué palabras claves tendrán mayor relevancia a la hora de clasificar en los buscadores la página web objeto de análisis.

Dado que no conocemos la forma en la que los motores de búsqueda analizan los sitios web y seleccionan las palabras que consideran clave, empleamos el resultado del análisis de forma meramente orientativa.

En concreto utilizaremos la herramienta Keyword Density Analyzer, de Seobook [\[69\]](#). Con esta herramienta podemos realizar un análisis pormenorizado del contenido de un sitio web, permitiéndonos dividir los resultados en función de la localización del contenido en el sitio web (cuerpo del texto, encabezados, links, imágenes)

- **Herramientas de pago**

Además de las herramientas anteriormente descritas, todas ellas gratuitas. Existen numerosas herramientas de pago, algunas con versiones de prueba (limitadas), especializadas en el proceso de investigación de palabras clave. Ofrecen funcionalidades similares a las anteriormente expuestas y algunas funcionalidades exclusivas propias de cada herramienta. Los datos de estas herramientas provienen de sus propias bases de datos, obtenidas de los propios ISP (compañías proveedoras de servicios de internet) y de su base de usuarios.

Algunas de estas herramientas son:

- Trellian's Keyword Discovery (<https://www.keyworddiscovery.com/>)
- Hitwise (<http://www.hitwise.com/gb/>)
- Searchmetrics suite (<https://suite.searchmetrics.com/en/research>)
- SEMrush (<https://www.semrush.com/>)
- WordStream (<https://www.wordstream.com/>)

Para ver el coste que supone el uso de estas herramientas tomaremos como ejemplo Trellian's Keyword Discovery. Existen distintos tipos de suscripción y de rango de precio: versión estándar (Gratuita, con limitaciones), versión Profesional (49,95\$/mes - 499,95\$/año) y por último una versión Enterprise (495\$/mes - 4752\$/año). Existe también una versión gratuita cuyas funcionalidades están limitadas.

El carácter de este proyecto no nos permite el acceso a las versiones de pago y por tanto se ha hecho uso de las herramientas gratuitas.

3.2.1.2.4. Proceso de investigación de palabras clave

Sobre estas líneas se ha intentado exponer, en líneas generales, las técnicas y herramientas empleadas en un proyecto SEO normal. En nuestro caso nos encontramos ante un proyecto atípico en cuanto a la investigación de palabras clave, y esto se debe a que el grueso del contenido del sitio web del Multiscale Materials Modeling Group, cerca del 90%, reside en las publicaciones del mismo, y estas no son objeto de modificación. Además, nos encontramos con que el lenguaje empleado es de carácter técnico, reduciendo significativamente las posibles modificaciones con el fin de adaptar el contenido a las necesidades del usuario (Ej. posibles sinónimos empleados con mayor frecuencia por los usuarios, empleo de lenguaje coloquial...). Paradójicamente, este lenguaje técnico simultáneamente dificulta y simplifica el proceso de investigación de palabras clave.

El hecho de que el lenguaje sea técnico nos limita en varios aspectos. En primer lugar, destacar la ausencia de sinónimos o alternativas a considerar para la gran mayoría de los términos empleados, lo que reduce notablemente las opciones. Por otro lado, muchos de los términos principales empleados en los contenidos de la página ni siquiera cuentan con datos de volumen de búsqueda, lo que denota la baja relevancia de dichos términos e imposibilita la evaluación de su relevancia relativa respecto al resto de palabras claves del sitio web. Paradójicamente esto hace que el proceso de investigación de palabras clave se vea simplificado.

Pese a estas restricciones se ha llevado a cabo un procedimiento similar adaptando las técnicas descritas y empleando las mismas herramientas que en un proyecto SEO normal. El objetivo es conocer las palabras clave que caracterizan las páginas del sitio web del proyecto, conocer el volumen de búsquedas que reciben dichos términos, observar el nivel de competencia por cada uno de los términos, encontrar posibles alternativas de mayor relevancia que los términos empleados hasta el momento, ... Por último, emplearemos la lista de palabras clave final del proceso para llevar a cabo un seguimiento del grado de indexación y el posicionamiento de las páginas del sitio web en los motores de búsqueda.

- Fase 1: Lápiz y papel

Adaptamos la técnica a nuestras necesidades. En nuestro caso nos limitaremos a analizar las palabras clave empleadas a lo largo del sitio web, tanto en el contenido como en la estructura del mismo.

En primer lugar, comenzaremos con la página principal, pues contiene la principal palabra clave para la que deseamos posicionar el sitio web. No es otra que *Multiscale materials modeling*. Incluimos también algunas de sus variaciones como *materials modeling* o *multiscale materials*.

Continuaremos el análisis con las secciones que alojan el grueso del contenido del sitio web, las secciones 'Research' y 'Publications'.

Comenzaremos analizando la sección 'Research'. Como podemos ver en la Figura 19, esta cuenta con una página principal y cuatro subsecciones, cada una con su propia página: *Atomistic modeling techniques*, *Irradiation of materials*, *Defects in semiconductors* y *Technological processes*. Evidentemente todas ellas forman parte de la lista de palabras clave pues en torno a ellas se desarrolla el contenido de esta sección.

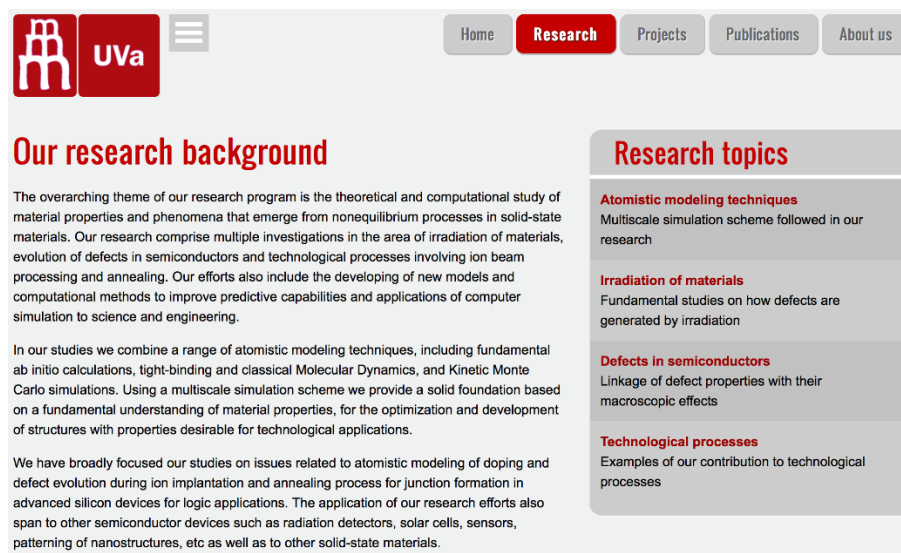


Figura 19. Página principal de la sección 'Research'.

Así mismo cada una de las subsecciones contará con una breve introducción y pasará a desarrollar los temas que trata (en la propia página de la subsección). Los títulos de los temas serán incluidos en la lista de palabras clave dado que representan el grueso del contenido de cada una de las subsecciones. En la Figura 20 tenemos un ejemplo en el que podemos ver los temas tratados en la subsección 'Atomistic modeling techniques'.

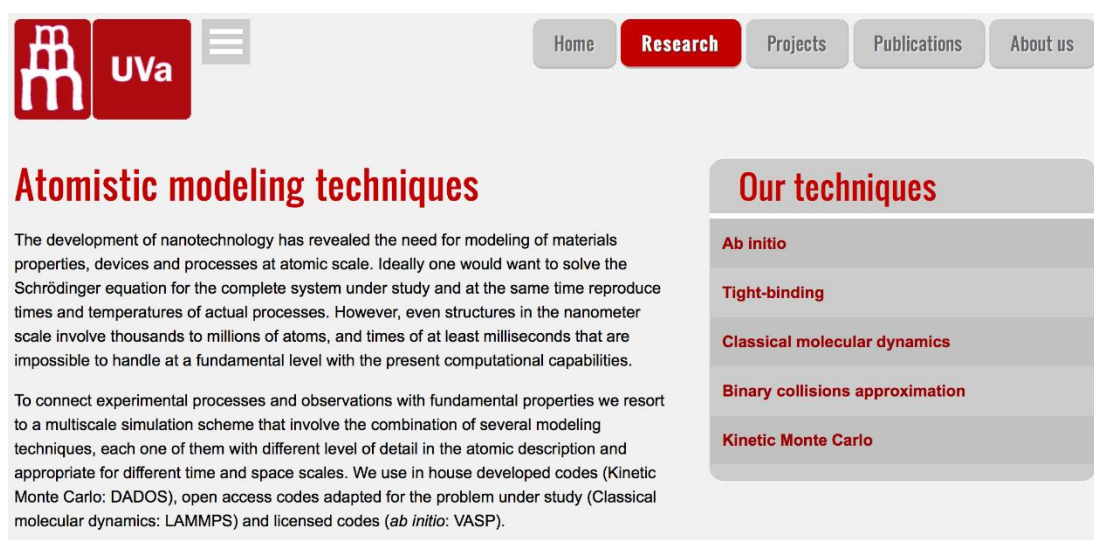


Figura 20. Subsección *Atomistic modeling techniques*.

La lista completa de los temas tratados en cada una de las subsecciones es la siguiente:

Atomistic modeling techniques

- *Ab initio*
- *Tight-binding*
- *Classical molecular dynamics*
- *Binary collision approximation*
- *Kinetic monte carlo*

Defects in semiconductors

- *Structural characterization*
- *Energetic characterization*
- *Electronic properties*
- *Dynamical studies*

Irradiation of materials

- *Damage generation mechanism*
- *Description of irradiation damage*
- *Molecular implants*
- *Influence of irradiation parameters*

Technological processes

- *Semiconductor doping*
- *Defect engineering*
- *Sputtering*
- *Nanodevices*
- *Thermal treatments*
- *Amorphous films*
- *Photoluminescence*

Las palabras claves obtenidas hasta ahora son las más obvias, pues son la base en torno a la que se construye los contenidos de la sección 'Research'. Para obtener las palabras clave que no son tan obvias analizaremos en detalle el contenido de cada una de las subsecciones y temas que acabamos de ver. Para ello haremos uso de la herramienta Keyword Density Analysis descrita previamente.

Como resultado obtendremos una lista con las palabras que se repiten con mayor frecuencia, de forma similar a como se muestra en la Figura 21.

KEYWORD DENSITY ANALYZER

Summary

Page title: Multiscale Materials Modeling Group @ UVA website :: Research :: Atomistic modeling techniques
Meta keywords: Multiscale materials modeling, atomistic simulation, modeling, front end processes, defect engineering, semiconductors, silicon, Si, germanium, Ge, amorphous silicon, amorphous germanium, binary collisions approximation, kinetic monte carlo, empirical potentials, tight-binding, ab initio, molecular dynamics, classical molecular dynamics, damage generation, damage, amorphous pocket, cluster, amorphous layer, defect cluster, point defect, interstitial, vacancy, bond defect, IV pair, diffusion, self-diffusion, dopant diffusion, diffusion path, migration energy, energy barrier, dopant, boron, B, boron interstitial cluster, BICs, fluorine, F, octodecaborane, energy gap, band diagram, defect levels, mid-gap levels, carrier trap, charge localization, carrier recombination, radiative recombination, ion implantation, plasma immersion implantation, molecular implant, cluster implant, cold implant, co-implantation, sputtering, irradiation, thermal annealing, flash annealing, spike annealing, laser annealing, rapid thermal annealing, solid phase epitaxial regrowth, SPER, nanodevice, FinFET, solar cell, radiation detector, photoluminescence, microscopy
Meta description: Atomistic simulation techniques used by the Multiscale Materials Modeling research group of the University of Valladolid.
Total word count: 996 (1570 including stop words)
Unique words: 474 (564 including stop words)

[All text](#) [Body text](#) [Headlines](#) [Links](#) [Images](#)

All text

Word	Count	Density	2 Word Phrases	Count	Density	3 Word Phrases	Count	Density
defect	14	0.89%	of the	23	2.93%	kinetic monte carlo	8	1.53%
molecular	13	0.83%	ab initio	10	1.27%	classical molecular dynamics	8	1.53%
dynamics	13	0.83%	molecular dynamics	10	1.27%	multiscale materials	6	1.15%
simulation	12	0.76%	in the	9	1.15%			

Figura 21. Keyword density analyzer.

Una vez obtenidas la lista de palabras pasaremos a su selección, pues no todas nos valen. Nos quedaremos con aquellas palabras que, dentro de la página correspondiente, tengan un mayor número de repeticiones, es decir tengan una densidad mayor dentro de la página. Esto no será así en todos los casos pues, si bien una palabra puede repetirse en muchas ocasiones esta puede no ser relevante como palabra clave (Ej. “of the” en la imagen anterior). Cuando hayamos obtenido todas las palabras que consideremos relevantes procederemos a eliminar aquellas palabras repetidas, es decir que aparezcan en subsecciones diferentes.

A continuación, realizaremos una tarea similar con el fin de determinar las palabras clave de cada una de las publicaciones. En este caso no recurriremos a una herramienta de análisis de densidad de palabras clave. Obtendremos las palabras claves de diversas fuentes relacionadas con cada una de las publicaciones, y entre ellas, las propias publicaciones.

En primer lugar, realizaremos un análisis de los títulos de las propias publicaciones pues estos incluyen palabras clave que tendrán mayor relevancia, dada su posición prominente, tanto para los buscadores como para los usuarios. Además, muchas de las publicaciones incluyen un pequeño índice de palabras clave denominado “keywords”, que indica de un breve vistazo la temática de la publicación.

Cada una de las publicaciones cuenta con un enlace a su archivo PDF correspondiente y otro enlace a la página web de la revista científica donde fue publicado. Lo que nos lleva a la fase 2, el análisis de competidores.

- Fase 2: Análisis de competidores

Dada la naturaleza y la temática del sitio web no existen grandes competidores, o al menos no se adaptan al concepto de competidor que todos tenemos. En la auditoría previa vimos que los principales competidores son Wikipedia, otros grupos de investigación, sitios web dedicados a conferencias, ... y

las páginas de web de las revistas científicas donde los artículos del grupo de investigación fueron publicados.

De hecho, son nuestros principales competidores, pues tratamos de posicionar el mismo contenido. La mayoría de estas páginas webs ofrecen una lista de keywords relacionadas con la temática del artículo. Tienen el fin de proporcionar a los usuarios una imagen global de la temática del artículo clasificar la temática de los artículos dentro de la propia web y a su vez clasificar la temática de los artículos dentro de la propia web. Emplearemos estos datos para completar la información obtenida en el análisis de las palabras clave de las publicaciones.

- Fase 3: Lista final

Una vez tengamos la lista con todas las palabras claves procederemos a obtener el volumen de búsquedas correspondiente con cada una de ellas, con el fin de determinar su relevancia para los usuarios de los motores de búsqueda.

Para ello introduciremos la lista de palabras en la herramienta gratuita Google AdWords Keyword Planner. Esta herramienta nos proporcionará para cada una de las palabras incluidas datos acerca de su volumen de búsquedas, competencia y nos sugerirá una cifra de cada a una campaña de pago por clic (PPC), ofreciéndonos una pista de la dificultad que puede suponer clasificar para dicha palabra clave.

Como resultado de este análisis tenemos un listado final de palabras clave que emplearemos para realizar el seguimiento sobre los cambios realizados a la página. Además, nos permite visualizar la forma en la que se utilizan las palabras clave en el sitio web y, en el caso de que su uso no sea eficiente, realizar los cambios pertinentes, tanto en la estructura como en el contenido. No se han incluido algunas palabras por resultar demasiado comunes y con el fin de mantener una lista de un tamaño manejable a la hora de llevar a cabo el seguimiento.

La lista está compuesta por 180 palabras y está ubicada en la Tabla A.2. del Anexo.

Se ha intentado mantener un equilibrio en el tamaño de las palabras clave con el fin de tener una muestra lo suficientemente amplia de cada una de ellas. Cabe recordar que cuanto más larga sea la palabra clave más específica será la búsqueda y más fácil resultará clasificar para dicho contenido.

En base a esto último las palabras clave se han dividido en función de su tamaño de la siguiente manera:

- 1 palabra: 60
- 2 palabras: 80
- 3 palabras: 36
- 4+ palabras: 4

Como podemos ver en la lista aquellas palabras clave de carácter general, y formadas por una única palabra, como 'ion', 'silicon', 'research', 'simulation'... reciben un gran número de visitas al mes. Por el contrario, a medida que las palabras clave aumentan su tamaño se reduce significativamente el número de visitas recibidas. Además, las palabras clave más especializadas, pertenecientes al

vocabulario técnico empleado en el sitio web, reciben un número de búsquedas mensuales muy reducido, llegando incluso a no contar con datos de visitas pues no llegan a las 10 búsquedas.

La lista contiene palabras claves cortas, pero en nuestro caso contamos con nuestra propia lista de palabras clave pertinentes al 'long tail'. Esta lista está formada por los títulos completos de las publicaciones. Se llevará a cabo un seguimiento periódico de esta lista con el fin de comparar el estado de indexación y las posiciones obtenidas por los archivos PDF y por las nuevas páginas web dedicadas a cada una de las publicaciones.

Cabe señalar que el análisis de las palabras clave relativas a las publicaciones han dado muestra de un posible caso de canibalización de palabras, pues múltiples artículos cuentan con palabras clave similares. El seguimiento de las posiciones de las páginas del sitio web para la lista de palabras clave que hemos obtenido arrojará más luz sobre el problema.

3.2.1.3. Uso de las palabras clave

En un proyecto SEO normal, una vez que se ha llevado a cabo el proceso de investigación de palabras clave, se procedería a analizar el uso que se da a esas palabras clave tanto en el contenido como en la estructura del sitio web, y en los casos en los que sea necesario llevar a cabo las modificaciones pertinentes para adaptar el contenido al vocabulario empleado por los usuarios en sus consultas. Además, se tendrían en cuenta las palabras clave resultantes de la investigación a la hora de crear nuevos contenidos.

En nuestro caso limitaremos el análisis del uso de dichas palabras clave a la estructura del sitio web, y son varios los motivos.

- La práctica totalidad del contenido que alberga el sitio web se corresponde con las publicaciones del grupo de investigación, y no son objeto de modificación.
- El contenido hace uso de un lenguaje técnico y especializado, dirigido a un público muy reducido y con los conocimientos necesarios para emplear dicho lenguaje.
- El uso de este lenguaje, y el público al que va dirigido, reduce significativamente, e incluso anula, las alternativas para llevar a cabo posibles modificaciones con el fin de adaptar el contenido al lenguaje empleado por los usuarios.
- De la misma forma, la creación de contenido está sujeta a la utilización del lenguaje técnico pertinente y no puede adaptarse a un público más general.

Los resultados de la investigación de palabras vienen a afianzar las conclusiones que se obtuvieron en el análisis preliminar, llevado a cabo en la auditoria previa. Con el fin de hacer un uso más eficiente de las palabras claves del sitio web, de cara a una mejora en el posicionamiento, se llevarán a cabo las siguientes modificaciones en la estructura del sitio web:

- Reestructuración completa de la sección 'Research', en la que se creará una página web para cada uno de los temas tratados en las diferentes subsecciones.
- Reestructuración de la sección 'Publications', en la que se creará una página web por cada una de las publicaciones. El objetivo es tener un mayor control en la forma en la que los motores de búsqueda interpretan y muestran el contenido en sus SERPs, frente a la que nos ofrecen los archivos PDF. Además, nos permite incluir nuevas funcionalidades.

- Reestructuración de la sección 'Projects', en la que se creará una página web para cada uno de los proyectos de la subsección 'Research projects', en la que se hace un breve relato de la temática del proyecto, y hacemos un mejor uso de las palabras clave.

Todos estos cambios se explicarán en profundidad en futuras secciones de este proyecto, cuando se haga frente a los cambios relativos a la arquitectura y al propio código HTML.

3.2.1.4. Contenido vertical

El uso de contenido vertical es bastante limitado. Por un lado, tenemos las imágenes que acompañan a los contenidos de cada uno de los temas de la sección 'Research'. Además, la página de inicio cuenta con 'slide' que muestra imágenes dando a conocer los distintos temas tratados en el sitio web.

Por otro lado, tenemos los artículos académicos, que también son un tipo de contenido vertical. Google Search, el buscador de carácter general, muestra resultados de Google Scholar, el buscador vertical de Google especializado en literatura académica. En la Figura 22 podemos observar un ejemplo, en él se muestra la posición preferencial con la que cuentan los resultados verticales de Google Scholar cuando aparecen en Google Search.



Figura 22. Resultados Google Scholar en las SERPs.

Más adelante, en la sección 3.2.3.8.2 profundizaremos en el funcionamiento de Google Scholar y explicaremos los elementos necesarios para una correcta indexación del contenido.

3.2.1.5. Gestión del contenido duplicado

Como hemos visto a la hora de valorar la unicidad del contenido, nos encontramos con un posible problema de duplicidad de contenido. En este caso de carácter externo, es decir, contenido existente en otros sitios web, lo que conlleva el riesgo de que sea considerado como un plagio, robo de contenido... y, por tanto, ser objeto de las penalizaciones asociadas al algoritmo Panda. Pero no parece que vaya a ser el caso.

Hasta el momento los artículos académicos alojados en el sitio web han convivido durante años, sin problema alguno, con las versiones (exactamente iguales) que presentan las páginas web de las revistas y plataformas científicas donde fueron publicados. Y no se trata de los únicos casos.

Como se puede ver en la Figura 23, correspondiente a la SERP de uno de los artículos académicos, en la Web conviven múltiples versiones de un mismo documento, y no siempre se considera que estas incumplan las directrices de contenido duplicado. Generalmente los motores de búsqueda tratan de ofrecer cierta diversidad tanto en el contenido como en las fuentes del mismo.

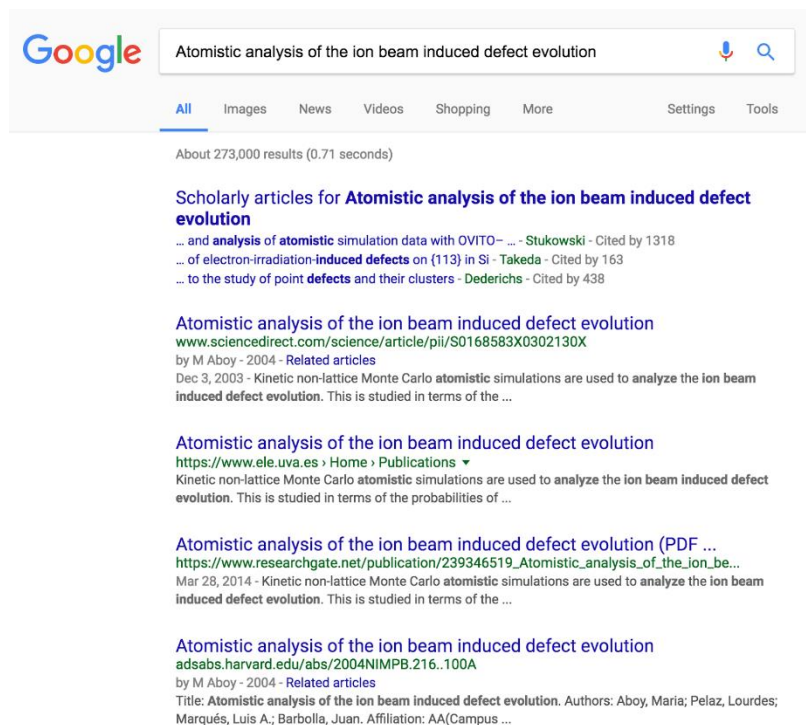


Figura 23. Múltiples versiones de un mismo artículo académico.

Una vez aclarado que en este caso las duplicaciones no tienen porqué acarrear penalizaciones, ahora el problema reside en la gran competencia que existe por clasificar los artículos académicos. En la Figura 23 se pueden ver hasta 4 versiones del mismo documento, pero puede haber muchas más. Llegado ese caso, los motores de búsqueda únicamente mostrarán aquellas que consideren relevantes, ocultando el resto. En nuestra experiencia en este proyecto, algunas de las publicaciones viven una situación intermedia, en la que periódicamente alternan las dos situaciones (visible y oculto), dificultando su asentamiento en las SERPs.

3.2.2. Arquitectura y aspectos técnicos

Una arquitectura óptima puede hacer que algunos de los procesos asociados a un proyecto SEO, como el rastreo y la indexación, puedan ver reducidos sus plazos sensiblemente. Por el contrario, una mala arquitectura, en la que existen trabas para que los motores de búsqueda realicen correctamente su trabajo, puede hacer que el proyecto SEO fracase. Incluimos algunos aspectos técnicos estrechamente con la arquitectura del sitio web.

En secciones previas de este documento, tras el análisis preliminar de la estructura del sitio web (sección 3.1.1), la investigación de palabras clave (sección 3.2.1.2) y cómo estas se emplean en el sitio web (sección 3.2.1.3), hemos llegado a la conclusión de que el sitio web necesita una reestructuración en algunas de sus secciones. Antes de comenzar con el proceso de reestructuración repasamos algunos de los factores que hacen que una estructura sea óptima tanto para los motores de búsqueda como para los usuarios humanos.

3.2.2.1. Arquitectura óptima

Cuando explicamos el funcionamiento básico de un motor de búsqueda, vimos que una de las etapas era el rastreo/crawling. En ella los robots de los motores de búsqueda se encargaban de recorrer los sitios web a través de sus enlaces, descubriendo nuevas páginas, tomando la información necesaria y creando/actualizando el índice. Por lo tanto, la existencia de una arquitectura óptima que facilite el rastreo de los sitios web es fundamental. Una arquitectura óptima es aquella que resulta fácilmente navegable tanto para los robots de los motores de búsqueda como para los usuarios.

Si por el contrario la arquitectura no es óptima, con páginas inaccesibles o empleando elementos que puedan dificultar la tarea de rastreo, podemos encontrarnos en la situación de que los motores de búsqueda no sean capaces de rastrear algunas de las páginas de nuestro sitio web. De esta forma cualquiera de las técnicas SEO aplicadas a lo largo de este proyecto resultarán inútiles para dichas páginas.

Por tanto, la arquitectura óptima de un sitio web deberá contar con dos aspectos fundamentales: accesible para los motores de búsqueda, de forma que facilite su trabajo, y una estructura interna óptima, bien categorizada y fácil de navegar tanto para los usuarios como para los motores de búsqueda. A continuación, profundizaremos sobre estos aspectos.

3.2.2.1.1. Accesible para los motores de búsqueda.

En este apartado veremos algunos los elementos que influyen en el grado de accesibilidad de un sitio web ante los motores de búsqueda. Se trata de aspecto relativos al tipo de contenido, elementos que dificultan el rastreo y elementos que los facilitan.

- Tipos de contenido

El contenido predilecto de los motores de búsqueda es aquel que se encuentra en archivos HTML. También son capaces de rastrear otros tipos de contenido, como imágenes y archivos Flash, pero su análisis resulta mucho más difícil y por tanto no resultan convenientes.

En el caso de las imágenes los motores de búsqueda tienen dificultad para establecer su relevancia pues cuentan con pocos elementos para establecerla. En concreto, cuentan con el nombre del archivo, su título y el contenido de la etiqueta *alt*.

- Arquitectura fácilmente rastreable

Como ya hemos visto, los robots de los motores de búsqueda emplean los enlaces para rastrear la Web. Por tanto, necesitamos que dichos enlaces sean fácilmente accesibles, de forma que no dificulten la tarea de rastreo llevada a cabo por los motores de búsqueda.

Existen una serie de elementos cuya utilización debemos evitar a la hora de crear la arquitectura de un sitio web [\[70\]](#). La mayoría de ellos dificultan, e incluso impiden, que los motores de búsqueda puedan acceder a los enlaces de las páginas web. A continuación, relatamos algunos de estos elementos:

- Enlaces en formularios: los motores de búsqueda no rellenan formularios, por tanto, todo el contenido que sea accesible únicamente a través de formularios será inaccesible para los robots de los motores de búsqueda.
- Enlaces en JavaScript: Hasta Junio de 2014 Google no procesaba los archivos JavaScript y por tanto los enlaces alojados en el código JavaScript eran inaccesibles para los robots. A partir de ese momento Google cuenta con la capacidad de procesar los archivos JavaScript y CSS, en algunos casos sólo de forma parcial. Por ello Google recomienda no bloquear el rastreo de este tipo de archivos [\[71\]](#).
- Enlaces en Java u otros plugins: son invisibles para los motores de búsquedas.
- Enlaces en Flash.
- Enlaces en archivos PDF o Powerpoint: Los motores de búsqueda son capaces de rastrear dichos enlaces y su peso será similar a los de un documento HTML. Pero existen problemas dependiendo del formato en el que fueran creado. Si cuentan con un formato basado de texto no deberían existir problemas. En cambio, si están basados en imágenes (como páginas escaneadas) los enlaces no podrán ser reconocidos por los robots de los motores de búsqueda.
- Enlaces a páginas bloqueadas con la etiqueta meta robots: les indica a los motores de búsqueda que no sigan (es decir, no rastreen) los enlaces salientes de dicha página.
(Ej.: `<meta name="robots" content="nofollow" />`)
- Enlaces con el atributo `rel="nofollow"`: Le indica al motor de búsqueda que no siga dicho enlace, es decir, que lo no rastree.
(Ej.: `Example`)
- Bloqueo con el archivo robots.txt: Desde el archivo robots.txt tenemos la capacidad de indicar al motor de búsqueda las páginas que no queremos que sean rastreadas.

Evitando el uso de estos elementos haremos que el sitio web sea mucha más fácil de rastrear, pues no existirán impedimentos para que los motores de búsqueda accedan a los enlaces.

En el caso que nos ocupa no se emplean ninguna de estas técnicas por lo que todos los enlaces del sitio web son completamente accesibles para los motores de búsqueda.

- Sitemaps

Sitemap es un protocolo que emplea el metalenguaje XML, y por el cual podemos proporcionar a los motores de búsqueda una lista con las URLs que deseamos que sean rastreadas e indexadas. Este protocolo está soportado por todos los grandes motores de búsqueda, entre ellos Google, Bing y Yahoo!.

Los sitemaps complementan el proceso de rastreo por parte de los motores de búsqueda, en ningún caso lo sustituyen. De hecho, la inclusión de una URL en el archivo Sitemap.xml no garantiza que dicha URL vaya a ser rastreada o indexada.

En palabras de Matt Cutts (Google) el hecho de que una página aparezca en el sitemap hace que esta tenga una oportunidad de ser rastreada e indexada. Ilustra esta situación con un ejemplo. La forma en la que Google rastrea e indexa las páginas web es a través de enlaces, no emplea para ello los sitemaps. Si por un casual Google logrará dar con la página A y la página B a través de enlaces, pero no con la página C, el hecho de que esta página aparezca en el archivo “sitemap.xml” hace que esta tenga la oportunidad de ser rastreada e indexada. Sin embargo, esto no supone la certeza de que esto vaya a ocurrir, simplemente se trata de una oportunidad que de otra forma no podría darse.

Los sitemaps proporcionan a los motores de búsqueda otro tipo de información, tal y como se puede ver en el ejemplo de la Figura 24:

- Fecha de la última modificación.
- Frecuencia de cambio de la URL.
- En el caso de la existencia de contenido duplicado los motores de búsqueda consultan el archivo sitemap para seleccionar la versión canónica.
- Registrar el archivo sitemap en herramienta como Google Search Console o Bing Webmaster Tools nos permite realizar un seguimiento sobre el estado de indexación de nuestras páginas.

Para elaborar un archivo “sitemap.xml” existen varias opciones:

- *Texto simple*. Bastará con crear un archivo con una URL por línea. Google recomienda que una vez tengamos este archivo utilicemos alguna de las herramientas de generación de sitemaps para adaptar dicha lista al protocolo XML Sitemap.
- *Generador de sitemaps*. Se trata de un programa que crea el archivo sitemap de forma automática a partir de una lista de URL (como en el caso anterior) o de una URL del sitio web, por conveniencia esta será la URL principal. En este último caso el programa recorrerá el sitio web y generará el sitemap con las URL que encuentre. Existen numerosas opciones donde elegir, en nuestro caso hemos empleado herramienta XML-Sitemaps Generator [\[72\]](#).

```

<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
    http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">
  <!-- created with Free Online Sitemap Generator www.xml-sitemaps.com -->

  <url>
    <loc>http://www.ele.uva.es/~mmm/</loc>
    <lastmod>2016-06-06T17:48:55+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>http://www.ele.uva.es/~mmm/research/</loc>
    <lastmod>2016-06-06T17:55:53+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>http://www.ele.uva.es/~mmm/projects/</loc>
    <lastmod>2016-06-06T17:53:04+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>http://www.ele.uva.es/~mmm/publications/</loc>
    <lastmod>2016-06-06T17:53:38+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>http://www.ele.uva.es/~mmm/contact.html</loc>
    <lastmod>2016-06-06T17:48:44+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>http://www.ele.uva.es/~mmm/research/atomistic-modeling-techniques/</loc>
    <lastmod>2016-06-06T17:49:13+00:00</lastmod>
    <changefreq>monthly</changefreq>
  </url>
  <url>
    <loc>http://www.ele.uva.es/~mmm/research/atomistic-modeling-techniques/ab-initio.
    html</loc>
    <lastmod>2016-06-06T17:49:10+00:00</lastmod>
  </url>

```

Figura 24. Ejemplo archivo sitemap.xml.

Hay que destacar que el tamaño máximo del archivo “sitemap.xml” es de 10 MB y que como máximo puede albergar 50.000 URLs. Si el sitio web con el que trabajamos supera estas cifras será necesario crear varios archivos sitemap y un archivo de índice de sitemap. Este archivo de índice también tiene limitaciones, su tamaño no debe superar los 10 MB y como puede albergar 1.000 sitemaps.

En el caso de que el sitio web cuente con páginas tanto en HTTP como en HTTPS solo será necesario incluir una de ellas en el sitemap.

Una vez generado el archivo “sitemap.xml” lo incluiremos en el directorio de mayor nivel de nuestro sitio web, el directorio raíz. En nuestro caso su localización es la siguiente: <http://www.ele.uva.es/~mmm/sitemap.xml>

El siguiente paso es hacer saber a los principales buscadores que nuestro sitemap ha cambiado. Para ello haremos uso de las herramientas para webmasters de cada una de ellas, es decir Google Search Console (Figura 25) y Bing Webmaster Tools (Figura 26). Ambas funcionan de forma similar, basta con proporcionar la ruta donde tenemos alojado nuestro sitemap y enviarlo.

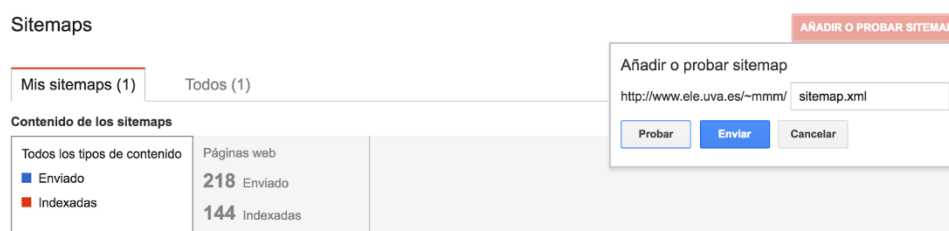


Figura 25. Funcionalidad sitemaps Google Search Console.

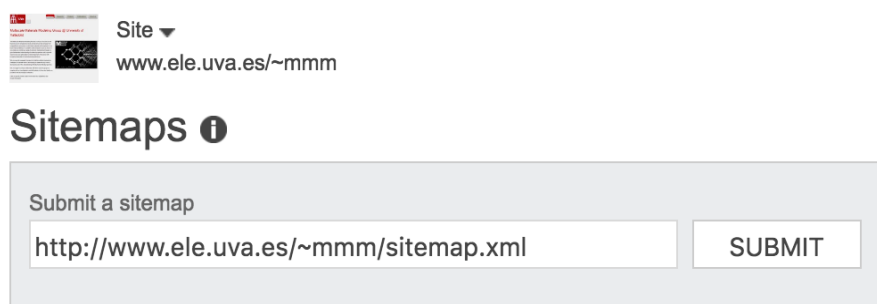


Figura 26. Funcionalidad sitemaps Bing Webmasters Tools.

A medida que el sitio web cambie, es decir, se incluyan nuevas páginas web, se modifiquen contenido de páginas ya existentes o bien se eliminen páginas, será necesario llevar a cabo un proceso de mantenimiento y actualización del archivo Sitemap.

3.2.2.1.2. Estructura interna óptima

A la hora de estructurar un sitio web hemos de plantearnos qué dificultades puede encontrar un usuario que visita la página por primera vez, pues dado que desconoce la estructura se sentirá perdido. Por ello es importante plantear una estructura que resulte reconocible por los usuarios, con elementos estándar empleados a lo largo de internet. Apoyada en una estructura de enlaces intuitiva de forma que la navegación por el sitio no se vea entorpecida. Una estructura así no solo ayuda a los usuarios, también a los motores de búsqueda.

En este apartado veremos algunos aspectos que deberemos tener en cuenta a la hora de estructurar un sitio web de forma óptima.

- Estructura lógica, basada en categorías

Emplear una estructura lógica, basada en categorías, es una cuestión fundamental pues nos permite que los usuarios puedan navegar a través del sitio web con mayor facilidad. Además, una estructura así también tiene beneficios para los motores de búsqueda, pues les permite conocer cómo está categorizado el contenido del sitio web y la jerarquía existente entre los distintos elementos de la misma. Los motores de búsqueda, con su experiencia en el rastreo de la Web, son capaces de detectar patrones en la estructura de los sitios web y tienen en mejor consideración a aquellos sitios que estructuran el contenido de forma intuitiva.

En nuestro caso, y dado que vamos a reestructurar algunas de las secciones del sitio web, haremos uso de una estructura basada en categorías, que nos permita separar los distintos contenidos y establecer una jerarquía entre los mismos. De esta forma permitiremos tanto a los usuarios como a los motores de búsqueda una navegación ágil en una estructura perfectamente reconocible en cualquier punto de la misma.

- Estructura plana vs. estructura profunda

A la hora de estructurar el contenido de un sitio web existen dos alternativas, la estructura plana y la profunda.

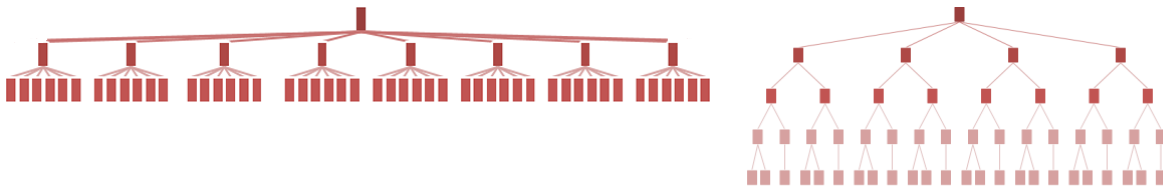


Figura 27. Estructura plana vs. estructura profunda.

Hay una regla básica a la hora de crear sitios web fáciles de utilizar, tanto por los motores de búsqueda como por los usuarios, y es mantener una estructura lo más plana posible. Una estructura plana requiere un reducido número de clics para acceder a cualquier página, mientras que las estructuras profundas requieren un mayor número de clics para acceder a las “páginas finales”, aquellas que generalmente contienen el contenido más detallado (Figura 27). Para cualquier sitio web con menos de 10.000 páginas, todo el contenido debe ser accesible a través de un máximo de 4 clics desde la página principal.

Desde el punto de vista del usuario una estructura plana resulta más simple que una profunda ya que se limita el número de páginas que ha de visitar para alcanzar el contenido deseado, reduciendo así la tasa de abandonos y aumentando las posibilidades de una futura visita.

Desde el punto de vista del motor de búsqueda, una estructura plana facilita el proceso de rastreo del sitio web, debido a que necesita seguir un menor número de enlaces para encontrar las “páginas finales”.

Aun así, cuando creamos estructuras planas debemos tener cuidado de no sobrecargar con demasiados enlaces. Una página web que cuente con un número excesivo de enlaces no podrá pasar mucho PageRank a las páginas que enlaza, diluyendo su relevancia. Por tanto, si bien resulta conveniente el uso de estructuras planas, no deben forzarse esta situación.

En nuestro caso, las necesidades que implican los cambios que vamos a aplicar en el proceso de reestructuración nos llevan a profundizar levemente en la estructura, con el fin de hacer un mejor uso del contenido y de sus palabras clave.

- Fácil de navegar

Como hemos visto una estructura lógica, basada en categorías, facilita la navegación por parte de los usuarios. A continuación, veremos algunos elementos que contribuyen a facilitar dicha navegación entre las distintas páginas del sitio web, y que permite a los motores de búsqueda a conocer mejor la estructura del mismo y a establecer las relaciones existentes entre las distintas páginas.

- *Breadcrumbs*: Los breadcrumbs, o migas de pan, nos permiten indicar a los usuarios el punto en el que se encuentran dentro de la jerarquía del sitio web, facilitando su navegación. Además, nos permiten reforzar la jerarquía interna de la estructura de enlaces mediante el uso de anchor text ricos en palabras clave. En la Figura 28 podemos ver un ejemplo. También tienen un componente estético y de usabilidad, pues aparecen en las SERPs, aunque para ello será necesario emplear un lenguaje de marcado de datos estructurado, como veremos más adelante.

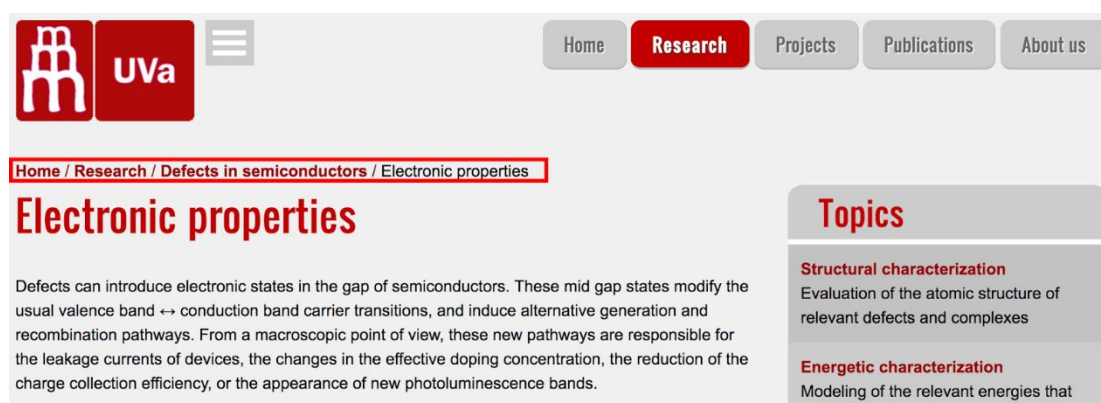


Figura 28. Ejemplo del uso de Breadcrumbs.

- **Contenido relacionado:** La incorporación de secciones de contenido relacionado dentro de las páginas web nos permiten acercar a los usuarios otros tipos de contenido, de forma que pertenezcan el mayor tiempo posible en la página web. Por lo tanto, se trata de un elemento que puede afectar a métricas de uso, como el tiempo de permanencia y la tasa de rebote, que afectan directamente a la posición de una página web.

Además, nos permiten reforzar la cohesión entre las distintas páginas web que cubren una misma temática. En nuestro caso hemos incluido una sección de “Publicaciones Relacionadas”, que nos permiten relacionar las publicaciones entre ellas, según su temática, y con otras secciones del sitio web.

3.2.2.1.3. Proceso de optimización de la arquitectura

Una vez que hemos visto los aspectos fundamentales que debemos tener en cuenta a la hora de crear una arquitectura óptima, los aplicaremos en el sitio web objeto de este proyecto.

En apartados anteriores, en concreto en la Sección 3.1, correspondiente con la Auditoria previa, y en la Sección 3.2.1.2, correspondiente con la Investigación de palabras clave, se consideró necesaria la reestructuración de las secciones ‘Projects’, ‘Publications’ y ‘Research’, especialmente esta última. El fin de esta reestructuración no es otro que hacer un uso más eficiente del contenido, y por consiguiente las palabras clave, con las que contamos.

Antes de comenzar a relatar los cambios efectuados en la estructura del sitio web veremos cuál es estado de la estructura del sitio web original.

- Estado inicial

Inicialmente el sitio web presenta una estructura plana, con una página principal y cuatro secciones principales. Como se puede ver en la Figura 29, cualquier contenido del sitio web se encuentra a un máximo de dos clics de la página principal.

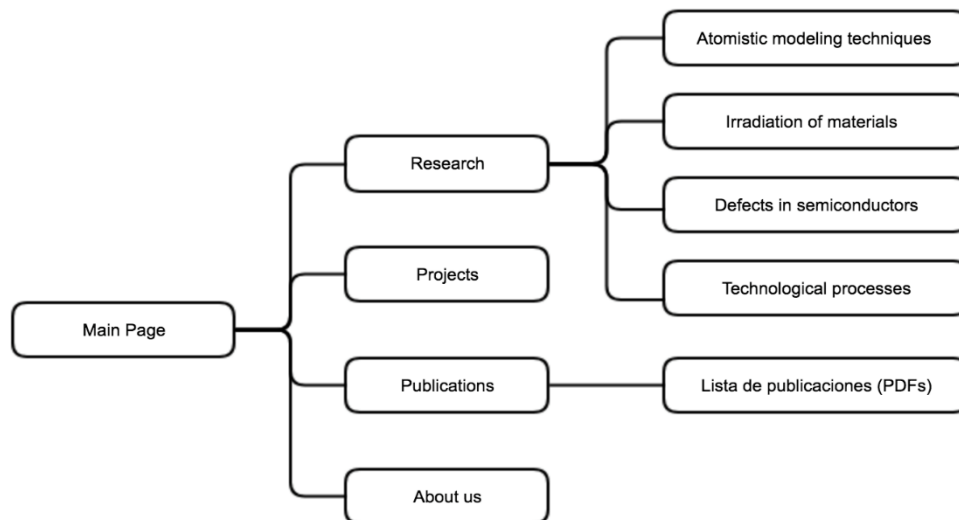


Figura 29. Estructura inicial del sitio web.

La reestructuración se llevará a cabo en términos similares, manteniendo una estructura plana, pero incluyendo cierta profundidad, necesaria cuando el objetivo es reestructurar el contenido, creando nuevas páginas web que traten ciertos contenidos de una forma mucho más específica, haciendo un uso más efectivo de las palabras clave.

Aunque en la imagen anterior podemos ver que, a través de la estructura de enlaces, se emplean categorías para clasificar el contenido, esto no ocurre a la hora de clasificar las páginas dentro del propio servidor. Todas las páginas se encuentran en el dominio raíz y por tanto la estructura de archivos no es coherente con la estructura de los enlaces. Esto se aprecia especialmente en las URLs. En el siguiente ejemplo podemos ver la diferencia entre la situación que relatamos y una correcta utilización de carpetas para categorizar el contenido.

<https://www.ele.uva.es/~mmm/research.processes.html#DefEng>

<https://www.ele.uva.es/~mmm/research/technological-processes/defect-engineering.html>

Este problema se debía quizás al bajo número de páginas, en concreto nueve. Y una estructura tan plana no compensaba el uso de carpetas para separar el contenido dentro del propio servidor. Dado que con la reestructuración se van a crear un número elevado de páginas la utilización de carpeta que categoricen el contenido es indispensable para mantener cierto orden interno, y ofrecer un mejor aspecto.

- Reestructuración de la sección 'Research'

Nuestro objetivo a la hora de reestructurar la sección 'Research' es hacer un uso más efectivo del contenido albergado en cada una de sus subsecciones. Inicialmente cada una de ellas trata varios temas diferentes, con su propio contenido (texto e imágenes), en una única página web.

Como vimos en la Auditoría previa esto supone un problema, pues salvo para un par de casos excepcionales, las páginas web de las subsecciones no aparecen en los resultados de búsqueda ante consultas que emplean las palabras clave de los diferentes temas que tratan. En cambio, si lo hacen cuando se trata de consultas relacionadas con los nombres de las propias subsecciones: *Atomistic modeling techniques*, *Irradiation of materials*, *Defects in semiconductors* y *Technological Processes*.

Como hemos visto, los motores de búsqueda examinan el contenido de las páginas con el objetivo de determinar su temática y las palabras clave para las que son relevantes. En este aspecto las páginas de las subsecciones están enfocadas, a través de sus títulos, encabezados..., a posicionar para sus propias palabras clave. Mientras, los contenidos de los distintos temas contribuyen a lograr esa posición, en lugar de tratar de posicionar para sus propias palabras clave, quedando diluidas entre todo el contenido.

Al dividir cada una de las subsecciones de la sección 'Research' con el fin de hacer un mejor uso de las palabras clave de las que disponemos, permitiremos que los motores de búsqueda asocien inequívocamente cada página con su contenido y de esta forma cada una de ellas clasificará para sus propias palabras clave.

Podría darse el caso de que las páginas de las subsecciones contarán con excelentes posiciones para las palabras clave a las que están dirigidas, y no compensará separar el contenido por temor a perder dichas posiciones, pero esto no es así. Durante la Auditoría pudimos ver que estas páginas cuentan con unos resultados más que discretos para las palabras clave a las que están dirigidas. Por tanto, no existe ningún impedimento para llevar a cabo la reestructuración del contenido, y crear nuevas páginas web para cada uno de los temas a tratar.

En la Figura 30 podemos ver la estructura final que presentará la sección 'Research' tras la reestructuración:

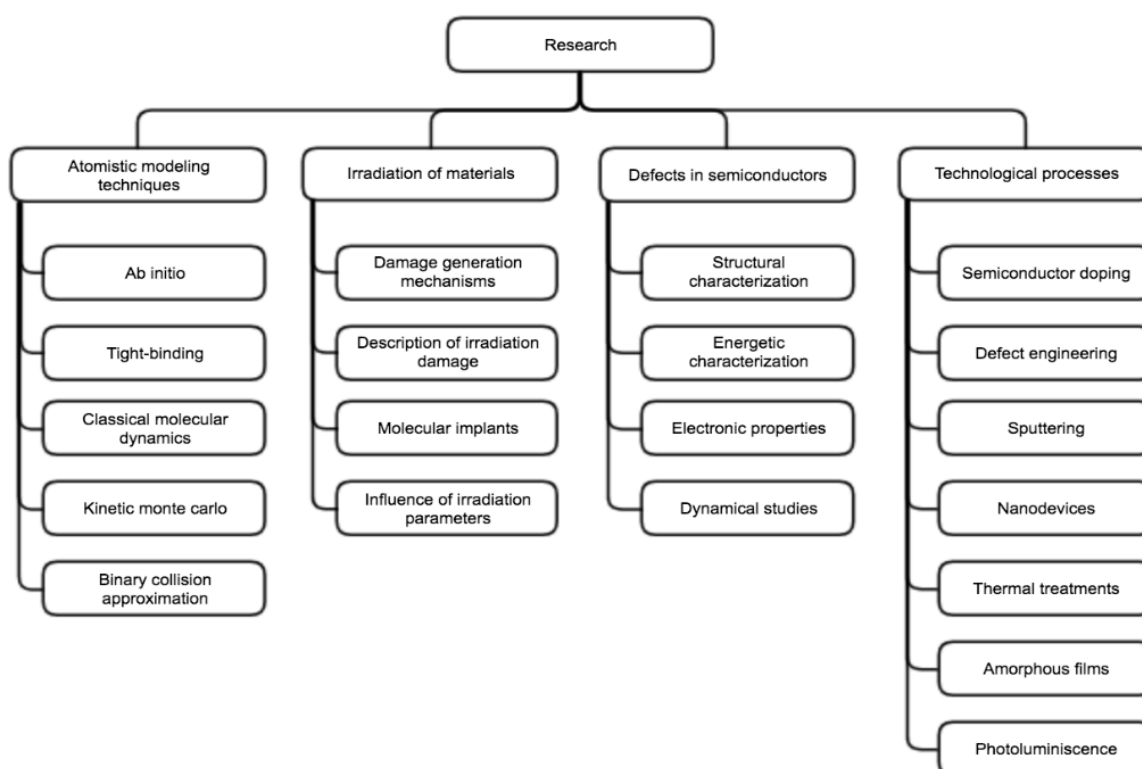


Figura 30. Reestructuración de la sección 'Research'.

Como podemos ver cada una de las nuevas páginas, correspondientes con cada uno de los temas, es enlazada desde la página de su subsección, manteniendo la jerarquía tanto en la estructura de

enlaces como en la estructura de archivos del servidor. Además, cada una de las páginas de una subsección enlaza con las demás, permitiendo navegar a lo largo de la subsección y logrando una mayor cohesión entre los contenidos a ojos de los motores de búsqueda.

- Reestructuración de la sección ‘Publications’

La sección ‘Publications’ cuenta originalmente, tal y como se puede ver en la Figura 31 imagen, con una única página web con la lista de las publicaciones llevadas a cabo por el grupo de investigación en las aparecen el título y los autores de cada una de ellas además de los enlaces a la revista científica donde fue publicada y al archivo PDF correspondiente.

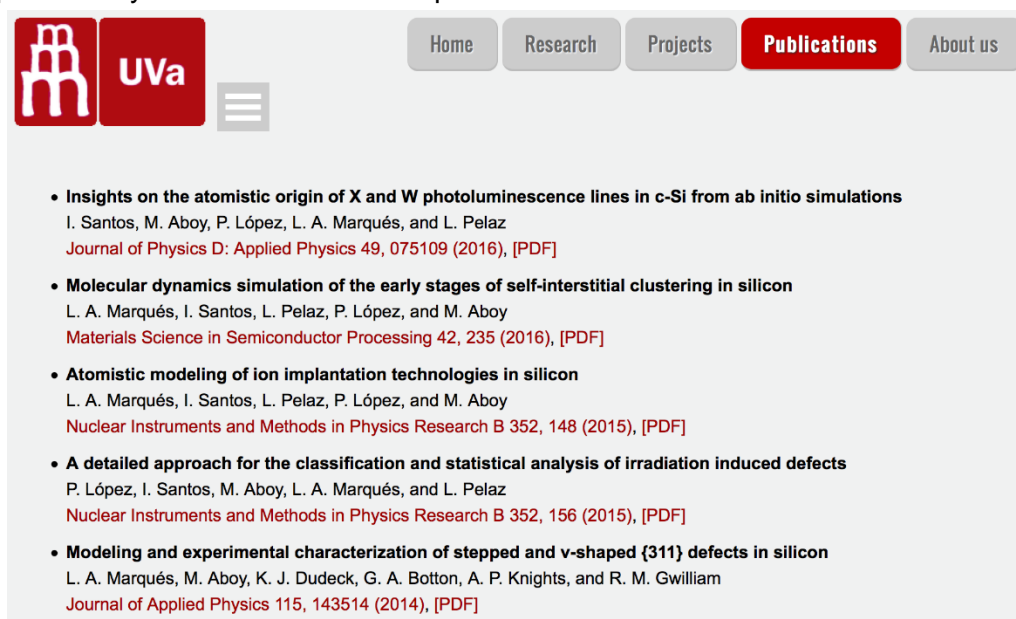


Figura 31. Situación inicial de la sección ‘Publications’.

Como vimos en la Sección 3.1, el estado de indexación de los archivos PDF correspondientes a las publicaciones nos lleva a realizar cambios en la sección ‘Publications’. Estos cambios se deben a que estas publicaciones suponen el grueso del contenido de la página y muchos de ellas no aparecen en el índice de Google o bien no lo hacen de forma estable. Además, nos encontramos con que las publicaciones que aparecen lo hacen con títulos y descripciones que no son los adecuados.

Creando una página para cada una de las publicaciones, en la aparezca el título, los autores, el nombre de la revista científica donde fue publicado, una breve descripción y los enlaces, tanto a la revista científica en la que fue publicado como al archivo PDF correspondiente, nos permitirán controlar mucho mejor la forma en la que aparecen en los buscadores. Además, nos permiten añadir funcionalidades que con los archivos PDF no podíamos contar, como los botones de redes sociales, secciones de Contenido relacionado... En la Figura 32 tenemos un ejemplo del formato que presentan estas nuevas páginas:

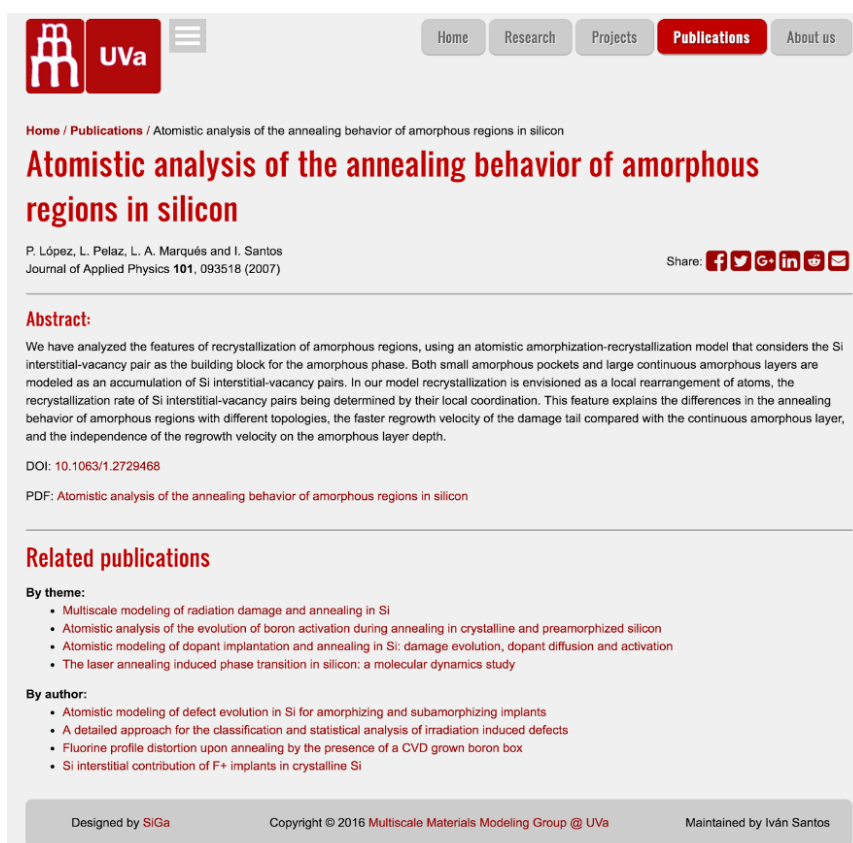


Figura 32. Ejemplo de las nuevas páginas web de las publicaciones.

El formato de la página principal en el que desde una única página se enlazan 90 publicaciones no es el más adecuado, pues el PageRank transmitido a cada una de las publicaciones será bajo. Para tratar de solucionarlo se añadirá una sección de “Publicaciones relacionadas”, tanto a las nuevas páginas de las publicaciones como a las páginas de la sección ‘Research’. De esta forma podremos relacionar las publicaciones que tengan una temática similar y también con las descripciones de las técnicas empleadas en las mismas, alojadas en la sección ‘Research’. En la imagen sobre estas líneas podemos ver un ejemplo del uso de las “Publicaciones relacionadas”, separados en función de su temática y su autor.

La estructura final de la sección ‘Publications’ tras la reestructuración presentará el aspecto reflejado en la Figura 33. En el que una página principal enlaza una serie de páginas web correspondientes con cada una de las publicaciones. Estas a su vez enlazan a su respectivo archivo PDF.

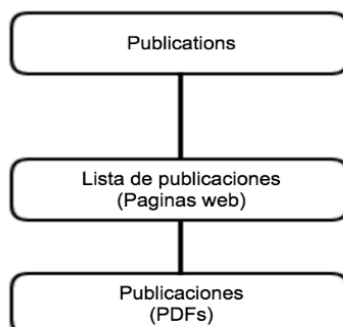


Figura 33. Reestructuración de la sección ‘Publications’.

- Reestructuración de la sección 'Projects'

En este caso se trata de un cambio menor, en el que se añaden algunas páginas con el fin de aumentar la oferta de contenidos del sitio web.

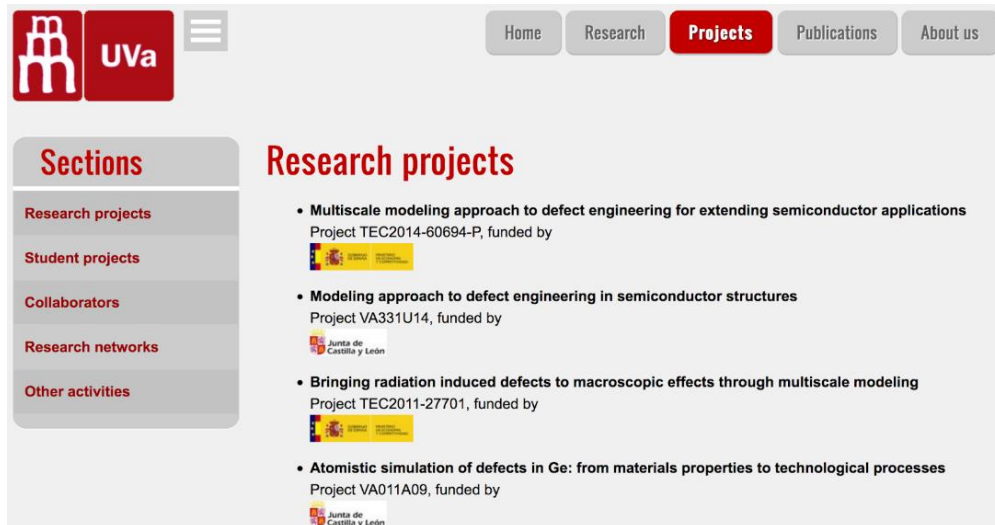


Figura 34. Situación inicial de la sección 'Projects'.

Tal y como podemos ver en la Figura 34, originalmente la sección 'Projects' cuenta con una única página web con varios apartados en los que se hace una relación de los Proyectos de investigación, Proyectos de estudiantes, Colaboradores, Redes de investigación asociadas y otras actividades en las que está involucrado el Multiscale Materials Modeling Group. En nuestro caso enfocaremos nuestra atención sobre el apartado 'Research Projects', que originalmente solo lleva a cabo una enumeración de los proyectos de investigación que están en marcha en ese momento.

De forma similar a como operamos en el caso de la sección 'Publications', crearemos una nueva página web para cada una de los 'Research Projects', con su título, información básica y una descripción del mismo. En la Figura 35 podemos ver un ejemplo.



Figura 35. Ejemplo de las nuevas páginas web de los 'Research Projects'.

Estas nuevas páginas se enlazan desde la página principal de la sección 'Projects' de forma que la nueva estructura de la sección presentará el aspecto reflejado en la Figura 36.

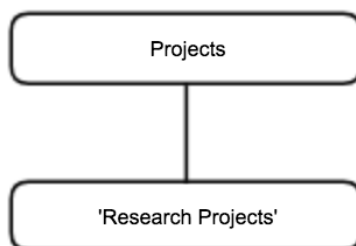


Figura 36. Reestructuración de la sección 'Projects'.

Estructura final

Una vez que se ha llevado a cabo la reestructuración de aquellas secciones que requerían cambios, la estructura final del sitio web es la siguiente (Figura 37):

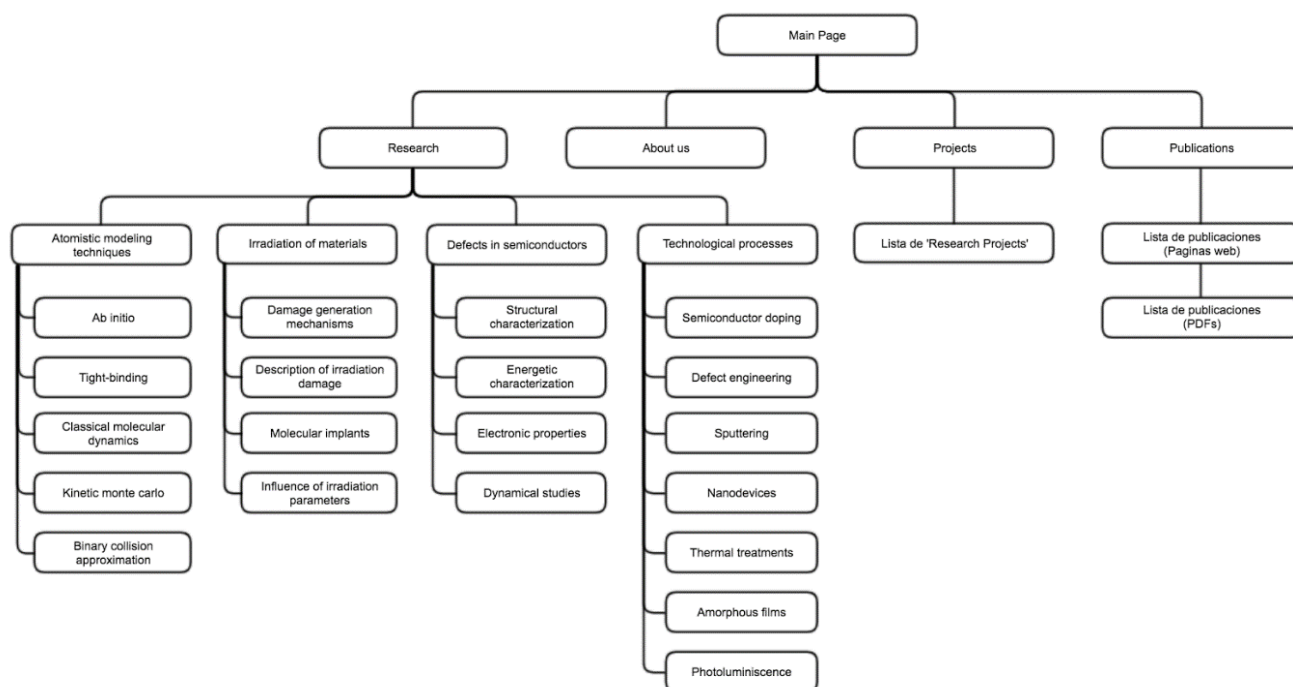


Figura 37. Estructura final del sitio web.

3.2.2.2. Formato de las URLs

El formato de las URLs no es, en sí mismo, un factor que afecte directamente a la posición de una página web. Pero si que lo hace de forma indirecta al estar estrechamente relacionado con otros factores. En primer lugar, cabe destacar que la práctica totalidad del formato de las URLs viene determinado por la estructura/arquitectura de archivos establecida en el sitio web y que, como hemos visto, resulta un factor determinante para facilitar el uso de la página tanto para los motores de búsqueda como para los usuarios.

Este último punto, los usuarios, nos permite abordar otro de los factores relacionados con el formato de las URLs, la usabilidad. Los motores de búsqueda emplean las métricas de uso (entre otros factores) para evaluar la usabilidad de un sitio web para finalmente, y junto con otros múltiples factores, determinar la relevancia de una página ante una consulta determinada.

Por tanto, las URLs deben ser descriptivas del contenido que albergan, comprensibles y de un tamaño que resulte fácil de leer y de recordar por los usuarios. Como ya hemos visto esto depende en gran parte del proceso de optimización de la arquitectura del sitio web realizado en la Sección 3.2.2.1, de hecho, algunas de las máximas aplicadas en ambos procesos coinciden.

A continuación, estableceremos unas pequeñas pautas a seguir a la hora de establecer el formato de las URLs:

- *URLs descriptivas:* Resulta conveniente que las URLs describan el contenido al que dirigen, coincidiendo en la medida de lo posible con los títulos de las páginas correspondientes y que, por tanto, hagan uso de sus palabras clave. Una URL que describa su contenido resultará más atractiva que a una que no lo haga.

<https://www.ele.uva.es/~mmm/research.mmm.html>

<https://www.ele.uva.es/~mmm/research/atomistic-modeling-techniques/>

Normalmente las URLs de los enlaces quedan ocultas tras los ‘anchor text’ o texto de enlace. Pero existen otras situaciones, como en las páginas de resultados, en las que las URLs aparecen como tal, y un formato descriptivo y el uso de palabras clave puede resultar determinantes.

- *URLs legibles y comprensibles:* Hacer uso de un lenguaje que resulte legible y fácil de entender resultará más atractivo para los usuarios. Puede marcar la diferencia entre obtener una visita o no. El siguiente caso lo ejemplifica claramente.

<https://www.ele.uva.es/~mmm/research/atomistic-modeling-techniques/>

<https://www.ele.uva.es/~mmm/2da321/amt934?HXID=324jqdna>

- *Tamaño de las URLs:* Las URLs más cortas son más fáciles de leer, analizar, copiar, pegar, compartir en redes sociales... No existe un tamaño estándar, pero la convención es que el tamaño no debe superar los 100 caracteres.
- *Evitar el uso de parámetros dinámicos:* Está directamente relacionado con la legibilidad y la compresión, pues generalmente la inclusión de parámetros la hace más difícil.
- *Uso de carpetas:* Es recomendable hacer uso de carpetas para jerarquizar el contenido, de esta forma los usuarios pueden comprender mejor la estructura del sitio web y su situación dentro del mismo. Pero su uso debe estar siempre dentro de lo razonable y ajustándose al contenido. Su uso viene marcado por estructura de archivos del sitio web.
- *Usar el guion (-) como separador:* A la hora de separar las distintas palabras de una URL resulta conveniente usar guiones (-). También se acepta el uso del signo de suma (+) y de la barra baja (_). Por otro lado, se desaconseja emplear espacios debido a que en las URLs son sustituidos por %20, afectando negativamente a la legibilidad.
- *Evitar el uso de mayúsculas:* Los usuarios están acostumbrados a escribir las URLs en minúsculas, por lo que la inclusión de mayúsculas puede llevarlos a confusiones. Por otro

lado, tenemos que algunos tipos de servidor no distinguen entre mayúsculas y minúsculas (servidores Microsoft/IIS) mientras que otros sí que lo hacen (servidores Linux/UNIX). Esto nos lleva a que, en el caso de un servidor Linux/UNIX, un error a la hora de tipografía al emplear las mayúsculas derivará en un error 404, pues dicha página web no existe.

A continuación, mostramos algunos ejemplos de cómo se han transformado las URLs. El cambio se nota especialmente en las cuatro subsecciones de la sección 'Research', que ven modificada su formato. Además, mostraremos algunos ejemplos de las URLs correspondientes a las páginas de nueva creación, pertenecientes a las secciones 'Research', 'Publications' y 'Projects'.

- Comparativa del formato de las URLs antiguas y las nuevas URLs, para la sección 'Research':

<https://www.ele.uva.es/~mmm/research.mmm.html>

<https://www.ele.uva.es/~mmm/research/atomistic-modeling-techniques/>

<https://www.ele.uva.es/~mmm/research.irradiation.html>

<https://www.ele.uva.es/~mmm/research/irradiation-of-materials/>

<https://www.ele.uva.es/~mmm/research.defects.html>

<https://www.ele.uva.es/~mmm/research/defects-in-semiconductors/>

<https://www.ele.uva.es/~mmm/research.processes.html>

<https://www.ele.uva.es/~mmm/research/technological-processes/>

- Ejemplos de nuevas páginas de la sección 'Research':

<https://www.ele.uva.es/~mmm/research/atomistic-modeling-techniques/ab-initio.html>

<https://www.ele.uva.es/~mmm/research/irradiation-of-materials/molecular-implants.html>

<https://www.ele.uva.es/~mmm/research/defects-in-semiconductors/dynamical-studies.html>

<https://www.ele.uva.es/~mmm/research/technological-processes/amorphous-films.html>

- Ejemplos de nuevas páginas de la sección 'Projects':

<https://www.ele.uva.es/~mmm/projects/>

<https://www.ele.uva.es/~mmm/projects/modeling-approach-defect-engineering-in-semiconductor-structures.html>

- Ejemplos de nuevas páginas de la sección 'Publications':

<https://www.ele.uva.es/~mmm/publications/>

<https://www.ele.uva.es/~mmm/publications/front-end-process-modeling-in-silicon.html>

<https://www.ele.uva.es/~mmm/publications/simulation-of-p-n-junctions.html>

<https://www.ele.uva.es/~mmm/publications/f-implants-in-crystalline-silicon.html>

<https://www.ele.uva.es/~mmm/publications/atomistic-simulation-techniques-in-front-end-processing.html>

<https://www.ele.uva.es/~mmm/publications/molecular-dynamics-simulation-early-stages-self-interstitial-clustering-silicon.html>

Como podemos ver se han seguido, en la medida de lo posible, las pautas establecidas. Se trata de URLs descriptivas (hacen uso de las palabras clave correspondientes), comprensibles, la adaptación de la nueva estructura nos permite ver el modo en el que el contenido está jerarquizado, ... En cuanto al tamaño, este se ha reducido todo lo que nos permite el principio de legibilidad. En algunos

casos, relativos a publicaciones con títulos especialmente largos, se ha sobrepasado el límite preestablecido de 100 caracteres en pos de la legibilidad.

3.2.2.3. Optimización para dispositivos móviles

A lo largo de los años hemos visto como la navegación desde dispositivos ha ido cobrando importancia hasta llegar superar, desde el año 2015, la navegación desde dispositivos de escritorio [49]. Dado que la experiencia de usuario en un dispositivo móvil es muy diferente de la de un dispositivo de escritorio, los motores de búsqueda decidieron incentivar la optimización para dispositivos móviles otorgando una pequeña mejora en la posición a aquellas páginas que estuvieran optimizadas frente a las que no lo estaban. Esta mejora solo afecta a las búsquedas realizadas desde dispositivos móviles.

Existen tres formas de enfocar la optimización para dispositivos móviles [73]. A continuación, realizaremos una breve descripción:

- *Sitio web para móviles* [74]: Este método emplea URLs independientes para las dos versiones de las páginas: móvil y de escritorio. Por tanto, se proporciona un código HTML diferente para cada uno de los dispositivos. Con este método se intenta detectar el dispositivo del usuario y posteriormente se le redirige a la versión correspondiente.

Este método requiere por tanto un especial cuidado, pues estamos duplicando el contenido del sitio web. Se hará uso de la etiqueta <link> junto con sus atributos *rel="canonical"* y *rel="alternate"*. La primera, situada en la versión móvil, indica la URL de la versión canónica de dicha página. La segunda, colocada en la versión de escritorio, indica la existencia de una versión alternativa, facilitando a los motores de búsqueda la detección de dichas versiones.

En cuanto a los redireccionamientos, se debe prestar especial atención para realizar redireccionamientos incorrectos. Además, resulta recomendable emplear la redirección 302.

- *Publicación dinámica* [75]: En este caso se emplea la misma URL para ambas versiones, pero se genera una versión diferente de código HTML para diferentes tipos de dispositivo, en función de la información recibida por el servidor sobre el navegador empleado por el usuario.

Con esta configuración se envía una respuesta desde el servidor con varios códigos HTML (y CSS) en una misma URL según el agente de usuario empleado en la solicitud. Para ello se emplea la cabecera HTTP Vary, que informa al navegador de que el contenido de la respuesta variará en función del agente de usuario que solicite la página.

- *Diseño adaptable* [76]: Se proporciona el mismo código HTML y la misma URL independientemente del dispositivo en el que navegue el usuario, y se modifica la presentación del contenido en función del tamaño de la pantalla del usuario. Para ello emplearemos el código CSS.

Google recomienda hacer uso de un diseño adaptable y estos son algunos de los motivos:

- Emplear una única URL facilita enlazar y compartir el contenido
- Al existir una única versión se facilita el rastreo e indexación de las páginas.
- Requiere menos tiempo de mantenimiento.
- Reduce la posibilidad de que se produzcan errores, como redirecciones erróneas.
- Se reduce el tiempo de carga pues no existe la necesidad de redireccionar a los usuarios.
- Los robots de los motores de búsqueda ahorran recursos a la hora de rastrear el sitio web, lo que puede derivar en que sean capaces de rastrear un mayor número de páginas y por tanto mejorar la eficiencia del rastreo.

En el caso que nos ocupa, el sitio web originalmente ya cuenta con un diseño adaptable por lo que poco se puede hacer en este aspecto, pues es el diseño que la propia Google recomienda.

3.2.2.4. Tiempo de carga.

El tiempo de carga es uno de los factores que afecta en el posicionamiento de una página web. De hecho, las páginas con un tiempo de carga menor tienen una pequeña ventaja frente a las páginas más lentas. Además, una página web rápida da buena imagen entre los usuarios y puede suponer la diferencia entre una visita o un rebote hacia la página de resultados.

Existen diversas herramientas que realizan un análisis del tiempo de carga de una página web, nos muestran que elementos ocupan en mayor o menor medida dicho tiempo y no indican qué acciones podemos llevar a cabo para reducirlo. Para realizar este análisis emplearemos las siguientes herramientas:

- *PageSpeed Insights* [\[77\]](#): proporcionada por Google y accesible desde Google Search Console. Muestra los resultados divididos en función del dispositivo en el que se muestra la página: móvil o escritorio. En el caso del móvil también nos proporciona información acerca de la experiencia de usuario de nuestra página en dichos dispositivos.
- *GTmetrix* [\[78\]](#): Esta herramienta gratuita realiza un análisis de la página web proporcionada por el usuario y nos muestra su puntuación para dos escalas diferentes (PageSpeed, Yslow), que tienen en consideración un conjunto de factores diferentes. En el caso de que alguno de estos factores afecte negativamente a la velocidad de carga la herramienta indicará qué elementos causan dichos problemas y ofrecerá posibles soluciones.

Hay que destacar que estas herramientas únicamente tienen en cuenta aspectos técnicos a la hora de realizar su evaluación. Por tanto, los cambios efectuados hasta este momento no deberían afectar a la velocidad de carga de la página.

Tras un primer análisis empleando la herramienta PageSpeed Insights, proporcionada por Google, nos encontramos con la situación reflejada por la Figura 38.

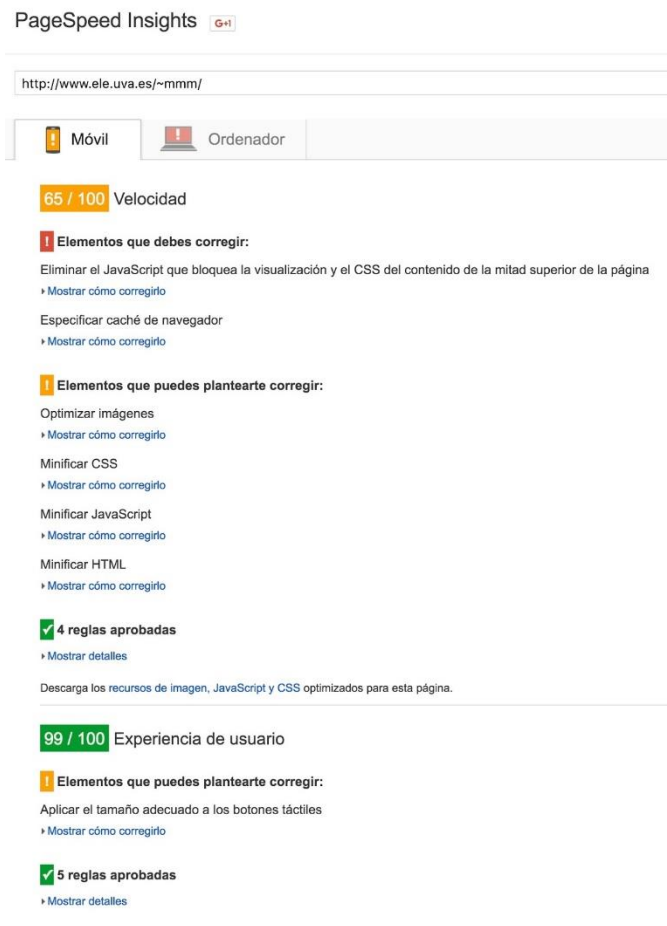


Figura 38. Situación inicial versión móvil. PageSpeed Insights.

Como se puede ver en la Figura 38, la herramienta otorga una puntuación en función de los requisitos que se cumplan, y nos detalla aquellos elementos que necesitan una corrección (obligatoria u opcional), tanto para dispositivos móviles como para dispositivos de escritorio. En nuestro caso los elementos a corregir son los siguientes:

- Elementos a corregir (obligatorios)
 - Especificar caché del navegador.
 - Eliminar el JavaScript que bloquea la visualización y el CSS del contenido de la mitad superior de la página.
- Elementos a corregir (opcionales)
 - Habilitar compresión. *
 - Optimizar las imágenes.
 - Minificar CSS.
 - Minificar HTML.
 - Minificar JavaScript.

**El elemento 'Habilitar compresión' no aparece en la imagen debido a que esta se tomó cuando el problema ya había sido subsanado. Se incluye en la lista porque formaba parte de la lista original de problemas y se procederá a explicar su corrección.*

En el caso de los dispositivos de escritorio, tal y como podemos ver en la Figura 39, los elementos a corregir son los mismos, solo que en este caso se altera la prioridad de dichos cambios. La optimización de imágenes cobra importancia en detrimento de la eliminación de los elementos que bloquean la visualización de la parte superior de la página.

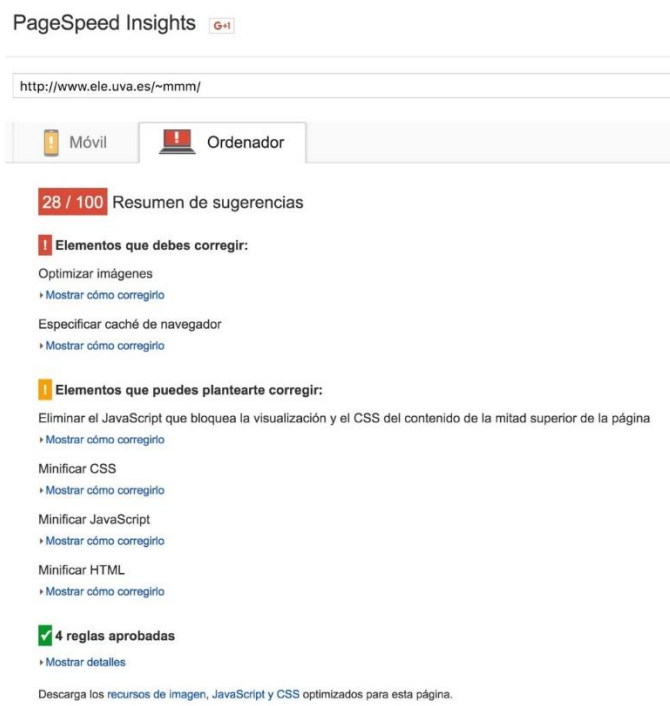


Figura 39. Situación inicial versión escritorio. PageSpeed Insights.

- Elementos a corregir (obligatorios)
 - Especificar caché del navegador.
 - Optimizar las imágenes.
- Elementos a corregir (opcionales)
 - Habilitar compresión. *
 - Eliminar el JavaScript que bloquea la visualización y el CSS del contenido de la mitad superior de la página.
 - Minificar CSS.
 - Minificar HTML.
 - Minificar JavaScript.

**El elemento 'Habilitar compresión' no aparece en la imagen debido a que esta se tomó cuando el problema ya había sido subsanado. Se incluye en la lista porque formaba parte de la lista original de problemas y se procederá a explicar su corrección.*

Una vez detectados los problemas que hacen que la velocidad de carga de nuestras páginas no tan rápida como deseamos pasemos a relatar sus soluciones. Destacar que, para cada una de las dos herramientas empleadas, tanto Google PageSpeed Insights como GTmetrix, nos proporcionan información para subsanarlos.

3.2.2.4.1. Habilitar la compresión

La mayoría de los servidores web tienen la capacidad de comprimir los archivos en formato gzip antes de enviarlos para su descarga por parte de los navegadores. Para ello pueden emplear módulos de terceros o rutinas incorporadas en el propio servidor. El objetivo de la compresión de estos archivos es reducir el tiempo necesario para descargar los archivos necesarios para visualizar un sitio web [79].

A continuación, veremos los pasos a seguir para habilitar la compresión en el servidor. En función del tipo de servidor será necesario emplear un procedimiento diferente. En el caso que nos ocupa tenemos un servidor Apache 2.4.7 sobre una distribución Ubuntu Linux.

Para habilitar la compresión en este tipo de servidor será necesario hacer uso del módulo Apache *mod_deflate* [80].

Como indica la propia documentación, el módulo *mod_deflate* nos proporciona el filtro de salida DEFLATE que permite que los archivos que salen de nuestro servidor sean comprimidos antes de ser enviados al cliente.

En principio el módulo *mod_deflate* viene incluido por defecto en la instalación de Apache, aun así, nos aseguraremos de que el módulo *mod_deflate* está instalado y habilitado en nuestro servidor Apache. Seguiremos los siguientes pasos:

- Comprobamos que el módulo *mod_deflate* está instalado. El módulo viene instalado por defecto en los servidores apache y se encuentra en la siguiente ruta: */usr/lib/apache2/modules/mod_deflate.so*.
- Comprobaremos que el módulo *mod_deflate* está habilitado, pero primero detallaremos la estructura del sistema de módulos. Los archivos de carga y configuración (archivos *.load* y *.conf*) de los módulos disponibles se encuentran en la carpeta */usr/lib/apache2/mods-available/*. En cambio, los módulos habilitados aparecen en la carpeta */usr/lib/apache2/mods-enabled/* como enlaces simbólicos a los archivos de la carpeta */usr/lib/apache2/mods-available/*. Por tanto, cuando activamos uno de los módulos se crea en la carpeta */usr/lib/apache2/mods-enabled/* un enlace simbólico a los archivos situados en */usr/lib/apache2/mods-available/*.

Una vez conocido el funcionamiento del sistema de módulos bastará con comprobar que los archivos de carga y configuración (*deflate.load* y *deflate.conf*) del módulo *mod_deflate* se encuentran en la carpeta */usr/lib/apache2/mods-enabled/*.

En nuestro caso el módulo estaba instalado, pero no habilitado, por lo que será necesario habilitarlo. Para ello existen dos alternativas:

- Mediante enlaces simbólicos, empleando el comando *ln -s*.
ln -s /etc/apache2/mods-available/deflate.load /etc/apache2/mods-enabled/deflate.load
ln -s /etc/apache2/mods-available/deflate.conf /etc/apache2/mods-enabled/deflate.conf

- Empleando el comando *a2enmod*, que habilita el módulo en cuestión de forma automática:
a2enmod deflate

Antes de terminar debemos asegurarnos de que la carga del módulo se realiza correctamente. Para ello comprobaremos en archivo de carga correspondiente, *deflate.load*. En él debe aparecer y permanecer accesible (no comentada) la siguiente línea de código:

```
LoadModule deflate_module /usr/lib/apache2/modules/mod_deflate.so
```

Por último, será necesario reiniciar el servidor apache para que los cambios efectuados cobren efecto, para ello emplearemos el siguiente código:

```
service apache2 restart  
apachectl -k restart
```

Para comprobar que la compresión ha sido activada existen varias opciones en nuestro caso emplearemos la herramienta 'HTTP Compression Test [\[81\]](#). Basta con incluir cualquier URL de nuestro sitio web y como resultado nos indicará si la compresión está activada.

Además, las herramientas sugeridas al comienzo de esta sección, Google PageSpeed Insights y GTmetrix, nos indicarán si la activación de la compresión se ha realizado correctamente, pues esta dejará aparecer como un problema a solucionar.

En el caso que nos ocupa el módulo estaba instalado, pero no habilitado. Se realizaron los enlaces simbólicos de forma manual, se comprobó que no había impedimentos para la carga de archivo y se reinició el servidor.

3.2.2.4.2. Especificar caché del navegador

Este problema surge cuando las respuestas del servidor no contienen cabeceras de caché de forma explícita o bien estas especifican un periodo de tiempo demasiado corto. En concreto Google considera que los recursos estáticos han de tener un ciclo de vida en caché de al menos una semana, mientras que para los recursos de terceros (anuncios, widgets...) ha de ser de al menos un día [\[82\]](#).

Empleamos el almacenamiento de recursos estáticos en la caché del navegador para ahorrar tiempos de carga a los usuarios que visitan el sitio web en más de una ocasión. Se aplicará a todo tipo de recursos estáticos como archivos CSS, JS, imágenes, ... En el caso de los archivos HTML dependerá de su código, pues en principio estos no se consideran como recursos estáticos no deberían almacenarse en caché por defecto.

Para establecer el ciclo de vida en caché de un recurso estático podemos emplear las siguientes cabeceras HTTP:

- *Expires*. Se utiliza para especificar el periodo de tiempo en el que el recurso está en caché. Durante ese periodo de tiempo el navegador utilizará el recurso en caché sin comprobar si existe una nueva versión en el servidor. Su formato es una fecha, y esta se corresponde

con el fin del periodo de tiempo que el recurso ha de permanecer en caché, como se puede ver en la Figura 40.

- *Cache-Control: max age*. Su funcionamiento es similar al de *Expires*. En este caso en lugar de una fecha almacena el número de segundos que dura el periodo de tiempo que el recurso permanecerá en caché.

Response Headers	
Date	Thu, 23 Jun 2016 14:52:34 GMT
Server	Apache/2.4.7 (Ubuntu)
Last-Modified	Wed, 22 Jun 2016 18:12:13 GMT
Etag	"3ad7-535ele0717940-gzip"
Accept-Ranges	bytes
Vary	Accept-Encoding
Content-Encoding	gzip
Cache-Control	max-age=2592000
Expires	Sat, 23 Jul 2016 14:52:34 GMT
Content-Length	4122
Keep-Alive	timeout=5, max=100
Connection	Keep-Alive
Content-Type	text/html
Request Headers	
Host	www.ele.uva.es
User-Agent	Mozilla/5.0 (X11; Linux x86_64; rv:45.0) Gecko/20100101 Firefox/45.0
Accept	text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language	en-US,en;q=0.5
Accept-Encoding	gzip, deflate
Connection	keep-alive

Figura 40. Ejemplo cabecera HTTP. *Expires* y *Cache-Control*.

En ambos casos cabe decir que estas cabeceras se aplicarán de forma incondicional. Una vez establecidas, y con el recurso descargado, el navegador no volverá a hacer peticiones de dicho recurso al servidor hasta que se alcance la fecha establecida en la cabecera *Expires*, o hasta que se borre la caché del navegador.

Expires tiene mayor compatibilidad que *Cache-control: max age* y por ello es preferible su uso. En ambos casos su valor debe ser de como mínimo una semana y como máximo un año. Superar el año infringe las normas RFC [\[83\]](#).

El sistema de cabeceras que acabamos de ver está automatizado, pero para ello será necesario activar el almacenamiento en caché del navegador en nuestro servidor.

De forma similar a como ocurrió al habilitar la compresión, será necesario habilitar el módulo Apache correspondiente, en este caso el módulo *mod_expires* [\[84\]](#).

Como indica la propia documentación, el módulo *mod_expires* está destinado a la generación de las cabeceras HTTP *Expires* y *Cache-Control* en las respuestas del servidor, de acuerdo con los criterios establecidos por el usuario. La fecha de expiración puede establecerse para que sea relativa al momento en el que el fichero de origen fue modificado por última vez, o bien al momento del acceso por parte del cliente.

El procedimiento será similar al caso 'Habilitar compresión'. Se llevarán a cabo las siguientes acciones:

- Nos aseguraremos de que el módulo se encuentra instalado. Por defecto se encuentra en la siguiente ruta: `/usr/lib/apache2/modules/mod_expires.so`.

- Comprobaremos que el módulo *mod_deflate* está habilitado. En el caso anterior, 'Habilitar compresión', se explicó la estructura del sistema de módulos, en este caso obviamos este paso. Se comprobará si el archivo de carga del módulo *mod_expires* (*expires.load*) se encuentra en la carpeta */usr/lib/apache2/mods-enabled/*. A diferencia del módulo de compresión, el módulo *mod_expires* no cuenta con archivo de configuración(.conf).

Nuevamente el módulo se encontraba instalado, pero no habilitado. Existen dos alternativas para habilitar un módulo:

- Realizar el enlace simbólico 'a mano' empleando el comando *ln -s*.
ln -s /etc/apache2/mods-available/expires.load /etc/apache2/mods-enabled/expires.load
- Emplear el comando *a2enmod*, que habilita el módulo en cuestión de forma automática:
a2enmod expires

Antes de terminar debemos asegurarnos de que la carga del módulo se realiza correctamente. Para ello comprobaremos en archivo de carga correspondiente, *deflate.load*. En él debe aparecer y permanecer accesible (no comentada) la siguiente línea de código:

```
LoadModule expires_module /usr/lib/apache2/modules/mod_expires.so
```

Por último, será necesario reiniciar el servidor apache para que los cambios efectuados cobren efecto, para ello emplearemos el siguiente código:

```
service apache2 restart  
apachectl -k restart
```

Hasta este momento únicamente hemos habilitado la posibilidad de especificar el tiempo en caché de los recursos desde el propio servidor. Para especificarlos será necesario añadir al archivo *.htaccess* el código de la Figura 41.

```
<IfModule mod_expires.c>  
# Enable expirations  
ExpiresActive On  
# Default directive  
ExpiresDefault "access plus 1 month"  
# My favicon  
ExpiresByType image/x-icon "access plus 1 year"  
# Images  
ExpiresByType image/gif "access plus 1 month"  
ExpiresByType image/png "access plus 1 month"  
ExpiresByType image/jpg "access plus 1 month"  
ExpiresByType image/jpeg "access plus 1 month"  
# CSS  
ExpiresByType text/css "access plus 1 month"  
# Javascript  
ExpiresByType application/javascript "access plus 1 year"  
</IfModule>
```

Figura 41. Código para la especificación del tiempo en caché.

En él se declaran los distintos tiempos de expiración en función del tipo de archivo (imagen, css, javascript) mediante la cláusula *ExpiresByType*. También es posible establecer un tiempo de expiración por defecto mediante la cláusula *ExpiresDefault*. En este ejemplo se establece tiempos de expiración relativos al último acceso al recurso, mediante el código *access*. También es posible establecer los tiempos de expiración en función de la fecha de modificación de los recursos, mediante el código *modification*.

Para comprobar que las cabeceras de almacenamiento en caché funcionan correctamente emplearemos las herramientas Google PageSpeed Insights, que fue la que reportó dicho error, y GTmetrix. Bastará con analizar cualquiera de las URLs del sitio web con dichas herramientas.

Si el proceso se ha realizado correctamente Google PageSpeed Insights deberá reportar que la necesidad de ‘Especificar caché del navegador’ ha dejado de ser un problema.

En cuanto a la herramienta GTmetrix, nos permitirá observar las cabeceras HTTP de las peticiones al servidor y comprobar que las cabeceras *Expires* y *Cache-Control* han sido añadidas, tal y como se ve en Figura 42.

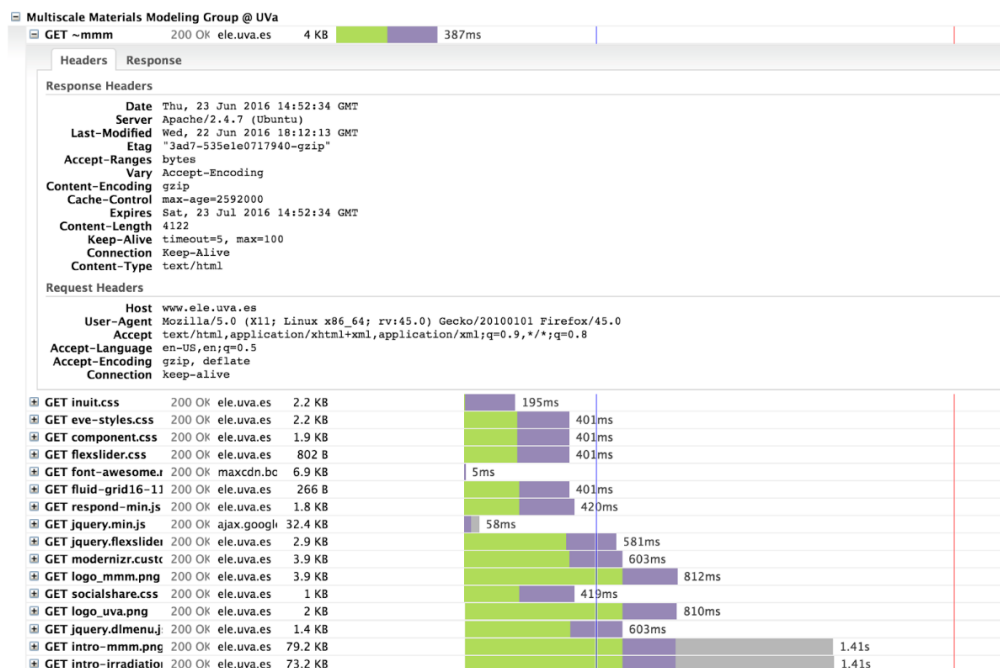


Figura 42. Cabeceras HTTP y carga de recursos. GTmetrix.

Hay que destacar que no todos los recursos son susceptibles de modificar el tiempo de expiración de su caché de navegador. En el caso del archivo JavaScript correspondiente a Google Analytics, el tiempo de expiración viene establecido por la propia Google y es de 2 horas, por lo que seguirá ocasionando problemas

3.2.2.4.3. Minificar CSS, HTML y JavaScript.

La minificación consiste en la eliminación de aquellos bytes que son innecesarios, como pueden ser espacios adicionales, comentarios, saltos de línea, sangrías... Al minimizar el código de estos recursos se reduce su tamaño y se acelera la descarga por parte del navegador, su análisis y el tiempo de carga [85].

La propia herramienta Google PageSpeed Insights nos ofrece la posibilidad de descargar versiones minimizadas de los archivos CSS y JavaScript. En el caso de los archivos HTML emplearemos la herramienta GTmetrix, que al igual que en el caso anterior ofrece una versión minimizada de los recursos HTML de la página web analizada. En la Figura 43 podemos ver una comparativa entre un archivo CSS y su versión minimizada:

```

/*-----BREADCRUMB.INUIT.CSS-----*/
/*
breadcrumb.inuit.css is an inuit.css igloo
igloos are CSS plugins which extend the inuit.css framework
They are released under the Apache License, Version 2.0 -- http:
@inuitcss
inuitcss.com
*/
/*-----USAGE-----*/
/*
Delete this section before go-live:

<ol class="nav breadcrumb">
  <li><a href="#">Home</a></li>
  <li><a href="#">Tutorials</a></li>
  <li><a href="#">CSS</a></li>
  <li>Creating a breadcrumb</li>
</ol>
*/

/*-----BREADCRUMB-----*/
/*
Create a breadcrumb trail navigation.
*/
.nav.breadcrumb li{
  font-weight:bold;
}
.nav.breadcrumb a{
  font-weight:normal;
}

/*
Here we set up the arrows using CSS generated content.
If you require this to work in IE7 uncomment the marked sections
*/
.nav.breadcrumb li:before{
  content:"\0020\003E\0020"; /* This is the CSS entity for ' »' */
}
.nav.breadcrumb li:first-child:before{
  content:normal;
}

```

Figura 43. Ejemplo minificación.

3.2.2.4.4. Optimización de las imágenes

Google PageSpeed Insights reporta este problema cuando detecta que las imágenes que contiene la página web objeto de análisis pueden ser optimizadas, reduciendo su tamaño sin afectar a su calidad.

Las imágenes pueden suponer una gran parte del peso de una página web y por lo tanto afectan a su tiempo de carga [86]. Por lo tanto, es conveniente llevar a cabo un proceso de optimización de dichas imágenes con el fin de reducir los tiempos de carga que sufren los usuarios. Además, y dado que buena parte de la navegación web se realiza desde plataformas móviles, esta reducción del tamaño de las imágenes puede suponer un ahorro en el consumo de los planes de datos de los usuarios.

Existen dos fases en la optimización de las imágenes:

- Optimización básica: consiste en cortar el espacio innecesario, reducir la profundidad de color hasta niveles aceptables, eliminar comentarios, reducir su tamaño (en píxeles) al establecido en la página, elegir el formato adecuado...
- Optimización avanzada: será necesario aplicar un proceso de compresión (sin pérdida) de los archivos PNG y JPEG.

Entre los elementos a optimizar en la opción básica se encuentra el formato. Es importante conocer que un mismo formato no es adecuado para todo tipo de imágenes. En nuestro caso, una página web, los mejores formatos son los siguientes:

- PNG: el formato general.
- GIF: adecuado para imágenes pequeñas (10x10 píxeles) o con una paleta de color de menos de 3 colores. Además, se emplean en archivos que contienen animaciones.
- JPEG: adecuado para imágenes de estilo fotográfico.

Para realizar las optimizaciones necesarias se ha utilizado la herramienta de edición fotográfica Photoshop, en concreto la utilidad “Guardar para Web...”, que nos permite seleccionar un formato adecuado para una página web, cambiar el tamaño de la imagen, reducir el número de colores... tal y como se puede ver en la Figura 44.

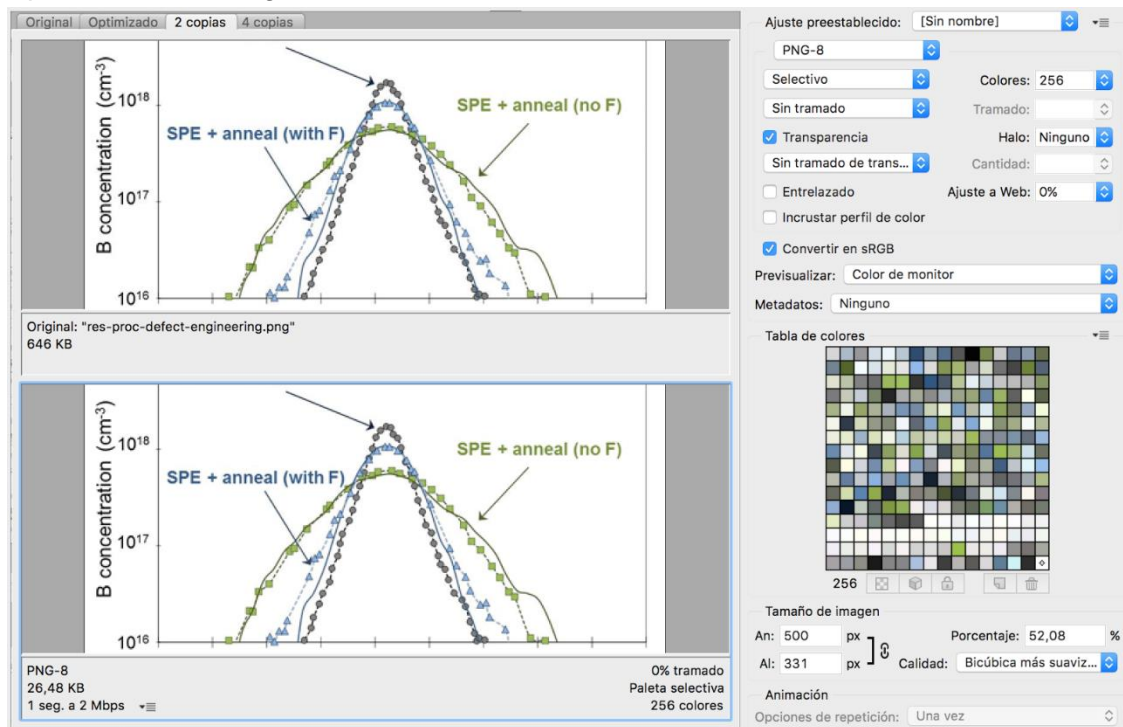


Figura 44. Optimización de una imagen.

La Tabla A.3 del Anexo nos muestra el ahorro en tamaño que ha supuesto el proceso de optimización. En cifras generales, ha supuesto un ahorro de más de 10 MB, un 62'87% menos que el tamaño original.

3.2.2.4.5. Eliminar el JavaScript que bloquea la visualización

Si durante la carga de una página por parte del navegador, este se encuentra con un script que bloquea la visualización del contenido de la página, el navegador ha de detenerse y descargar el archivo causante del bloqueo. Cada vez que esto sucede conlleva un retraso en la carga de la página. Existen varios métodos para tratar de evitar este problema [87].

Si un archivo JavaScript es necesario para la visualización del contenido de la parte superior de la página, trataremos de incluirlo en el código HTML. Este método es válido para aquellos scripts de tamaño reducido, como el ejemplo mostrado en la Figura 45.

```
<script type="text/javascript">
$(window).load(function() {
  $('flexslider').flexslider({
    animation: "slide",<!--you can also choose fade here-->
    directionNav: false,<!--Attention: if you choose true here, the nav-buttons will also appear in the ticker! -->
    keyboardNav: true,
    mousewheel: true
  });
});
</script>
```

Figura 45. Código JavaScript incluido en código HTML.

En cuanto a los archivos JavaScripts que añaden funcionalidades, estos deberán ejecutarse una vez el contenido haya sido mostrado. Para ello emplearemos la carga asíncrona.

La carga asíncrona consiste en ejecutar el archivo JavaScript tan pronto como sea posible sin detener la carga de la página. De esta forma la ejecución del archivo JavaScript se produce de forma asíncrona respecto al resto de la página. Para declarar la carga asíncrona de un archivo JavaScript será necesario emplear el atributo HTML *async* al cargar el script, como en el ejemplo de la Figura 46. Destacar que el atributo *async* solo es válido para scripts externos.

```
<script async src="/~mmm/js/respond-min.js" type="text/javascript"></script>
<script async src="/~mmm/js/modernizr.custom.js"></script>
<script async src="/~mmm/js/jquery.flexslider-min.js" type="text/javascript"></script>
```

Figura 46. Ejemplo de uso del atributo *async*.

Si los archivos JavaScript usan *document.write* no es seguro emplear la carga asíncrona. Además, será necesario prestar especial atención a las dependencias entre recursos a la hora de realizar cargas asíncronas.

Dado que en nuestro caso la carga asíncrona nos ocasiona problemas de compatibilidad entre los distintos archivos JavaScript, se empleará otro método para evitar que los archivos JavaScript bloqueen la visualización de la página. El método empleado no es otro que retrasar la carga de dichos archivos lo máximo posible, para ello bastará con colocarlos al final del cuerpo (<body>) del archivo HTML, en lugar de hacerlo en la cabecera.

Sin embargo, esta solución no es aplicable a todos los archivos JavaScript. El archivo *modernizr.custom.js* deberá permanecer en la cabecera, pues en determinados navegadores su ausencia puede provocar errores.

3.2.2.4.6. Optimizar la carga de CSS

Los CSS provocan un aumento en el tiempo de carga del contenido de una página, pues los navegadores se bloquean para su descarga antes de que apliquen el formato al contenido.

Las soluciones a este problema pasan por insertar en los archivos HTML el código CSS necesario para no demorar la aplicación del formato al contenido. En el caso de los archivos CSS pequeños será posible incluirlos totalmente en el código HTML [\[88\]](#).

En el caso de los archivos CSS grandes será necesarios localizar las partes necesarias para mostrar la parte superior de la página web, e insertarlas en el archivo HTML. En este caso el archivo CSS original, al que pertenecen los fragmentos de código, se cargará después de que se haya cargado la página. También es conveniente reducir el número de archivos CSS, unificando dichos archivos, con el fin de realizar menos llamadas al servidor.

En nuestro caso hemos agrupado el código de los distintos archivos CSS con los que contaba el sitio web originalmente en dos archivos, dependiendo de la función de dicho código CSS. Por un lado, el estilo general de la página, y por otro los estilos de las funcionalidades como el menú desplegable, el slider de la página principal, los breadcrumbs y los botones de las redes sociales.

También se han llevado a cabo pruebas colocando distintas porciones de código CSS en los archivos HTML con el fin de solucionar el problema, pero en ninguno de los casos la herramienta

Google PageSpeed Insights ha considerado que el problema estuviera solucionado. Por ello, finalmente, se ha decidido no colocar ningún tipo de código CSS en los archivos HTML.

Tras todos los cambios realizados nos encontramos que la puntuación ofrecida por Google PageSpeed Insights ha mejorado tanto en la versión móvil (Figura 47) como en la de escritorio (Figura 48), especialmente en el caso de esta última. Sigue habiendo elementos que causan problemas, pero como hemos visto estos no parece que tengan solución.

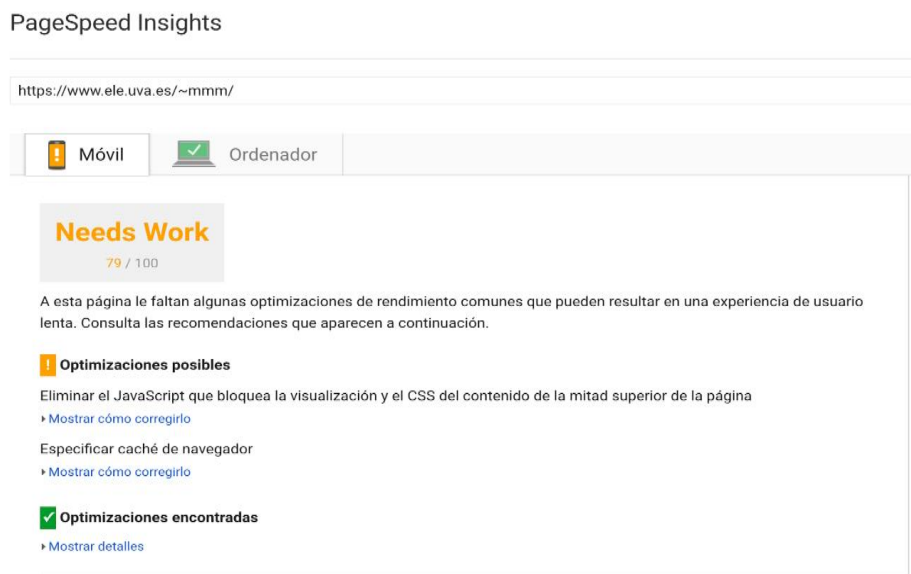


Figura 47. Situación final versión móvil. PageSpeed Insights.

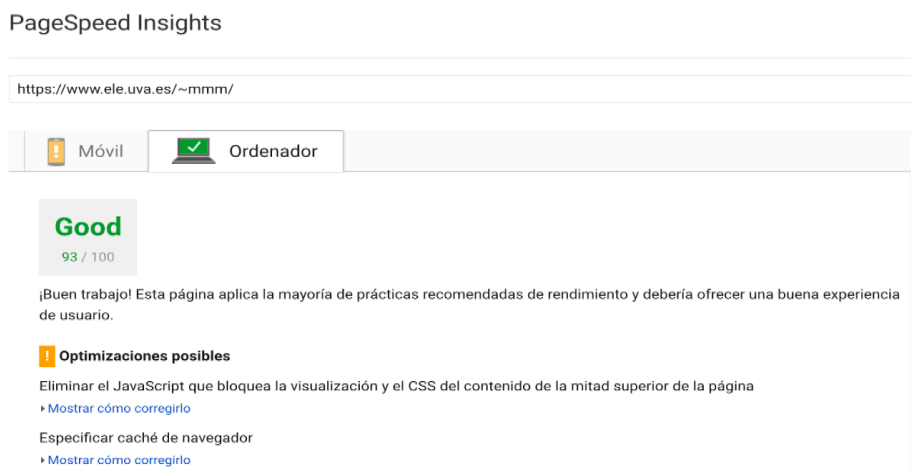


Figura 48. Situación final versión escritorio. PageSpeed Insights.

3.2.2.5. HTTPS/SSL

En Agosto del año 2014 Google anunció que, en pos de la seguridad, comenzaría a considerar el uso de HTTPS como uno de los factores que afectan a la posición de las páginas web dentro del ranking. En palabras de Google, el impacto de dicha actualización fue menor, afectando únicamente al 1% de las consultas, pero a medida que pase el tiempo este factor verá incrementado su peso, pues el objetivo final es incentivar a todos los propietarios de sitios web al cambio a HTTPS, de forma que la navegación en la Web sea más segura [35].

Existen numerosas alternativas a la hora de encriptar la información de un sitio web mediante la adquisición de certificados. Existen opciones de pago, como Symantec, Comodo, GoDaddy, GlobalSign, DigiCert, StartCom... cuyos precios van desde los 60€/año hasta 1.300€/año, dependiendo de la clase del certificado, el número de dominios sobre los que se puede aplicar, funcionalidades... Por otro lado, tenemos las opciones gratuitas, como StartCom y Let's Encrypt. En el caso de StartCom, mencionado anteriormente como una de las opciones de pago, resulta obvio que las prestaciones de la versión gratuita se ven reducidas frente a las de la versión de pago.

En nuestro caso, y dado que el sitio web pertenece a una institución educativa afiliada a la RedIRIS (Red española para Interconexión de los Recursos InformáticoS), haremos uso del servicio de certificación gratuita que esta ofrece a sus afiliados. La RedIRIS presta este servicio gracias a los acuerdos negociados a escala europea entre GÉANT (la asociación de redes académicas europeas, conocida anteriormente como TERENA), de la que es miembro, y los diversos proveedores de servicios de certificación

GÉANT tiene como objetivo “promover y desarrollar infraestructuras de redes internacionales de alta calidad para apoyar la investigación y la educación europeas”. La asociación ofrece otros servicios como eduroam, servicios en la nube...

Desde el año 2009 y hasta el año 2015, el proveedor de los certificados era Comodo. Una vez finalizado el contrato entre Comodo y GÉANT, se lanzó un concurso público en el que la empresa escogida como proveedor de los certificados fue DigiCert. Como podemos ver se trata de empresas líderes en el sector, por lo que la autoridad y calidad de los certificados está fuera de toda duda.

La puesta en marcha del certificado fue solicitada a los administradores del correspondiente servidor de la Universidad de Valladolid, y ellos mismos la llevaron a cabo. Una vez puesto en marcha el certificado es necesario realizar una migración del sitio web de HTTP a HTTPS. Cuando esto sucede Google considera que se está realizando el traslado del sitio web, con cambios en las URLs, como si de un nuevo dominio se tratara. Dado que Google trata a las páginas HTTP y HTTPS de forma separada, será necesario registrar una nueva propiedad en Google Search Console con la versión HTTPS del sitio web.

En nuestro caso, dado que la migración es completa, llevamos a cabo la redirección de todas las páginas web en HTTP a sus versiones HTTPS. Para ello haremos uso de la redirección 301, es decir, la redirección permanente. Se recomienda que esta redirección se lleve a cabo en el propio servidor y no a través del documento `.htaccess`.

3.2.2.6. Gestión del contenido duplicado

En la Sección 3.2.1.5 analizamos la gestión del contenido duplicado de carácter externo. Esta ocasión analizaremos el contenido duplicado interno, inherente al sitio web.

En nuestro caso, el sitio web del Multiscale Materials Modeling Group (<http://www.ele.uva.es/~mmm>), nos encontramos con que no existen elementos duplicados, ya sean páginas web o archivos PDF. En principio esto bastaría, pero por desgracia esto no es así.

El sitio web del grupo de investigación está alojado en el dominio del Departamento de Electricidad y Electrónica de la Universidad de Valladolid (<http://www.ele.uva.es>), y es solo una pequeña parte del mismo. En dicho dominio se alojan otros muchos sitios web correspondientes a otros grupos de investigación, a páginas personales de profesores e investigadores miembros del departamento... De hecho, algunos de los miembros del Multiscale Materials Modeling Group cuentan con su propio sitio web. Esto no supondría un problema si no fuera por el hecho de que dichos sitios web alojan exactamente el mismo contenido que nuestro sitio web, ya sean archivos PDF o páginas web de publicaciones.

Al encontrarse todo este contenido bajo el mismo dominio, cuando los motores de búsqueda se detectan varias versiones de un mismo documento elegirán una de ellas, la que consideren más relevante, la mostrarán en las páginas de resultados y ocultarán el resto.

Cuando llegue el momento de la visualización de los resultados mostraremos algunos casos de palabras clave que alternan periodos en los aparecen y desaparecen de los resultados de búsqueda. Pero se ha comprobado que estas no desaparecían, simplemente eran reemplazadas por una versión del mismo documento alojada en otro de los sitios web alojados en el dominio.

La solución a este problema, en el caso de los archivos PDF, pasa por alojar todos los archivos en un mismo lugar a modo de repositorio y enlazar dichos archivos desde todas las páginas. En el caso de las páginas de publicaciones duplicadas, bastaría con incluir una etiqueta `<link>` con el atributo `rel="canonical"` en aquella versión que consideremos que es la duplicada indicando la URL de la versión canónica u original, tal y como podemos ver en el ejemplo.

```
<link rel="canonical" href="http://www.ele.uva.es/example">
```

También podemos hacer uso de redirección 301 es decir, permanentes, pero en este caso es una solución demasiado drástica.

3.2.3. Código HTML

En este apartado abordaremos las modificaciones llevadas a cabo sobre los principales elementos del código HTML que los motores de búsqueda consideran a la hora de establecer la posición de una página web. La mayoría de estos elementos se emplean para determinar la temática de la página web, tanto por los motores de búsqueda como por los usuarios.

Por tanto, resultará conveniente hacer un uso más eficiente de las palabras clave más significativas de nuestro sitio web, empleándolas en dichos elementos HTML, de forma que los motores de búsqueda puedan reconocer con mayor facilidad la temática de las páginas y estas obtengan mejores posiciones para las consultas relacionadas con dichas palabras clave.

Estos elementos HTML también tienen asociado un componente de usabilidad. Cuando un usuario lleva a cabo una búsqueda orgánica elementos como el título o la descripción, suponen el primer punto de contacto con el usuario y pueden determinar la diferencia entre recibir una visita o no recibirla. Esto

afecta a las métricas de uso del sitio web y, por tanto, termina afectando a la posición de las páginas en los rankings.

3.2.3.1. Etiquetas de título

La etiqueta de título (<title>) es el elemento más importante dentro del código HTML. Como ya hemos comentado, los motores de búsqueda emplean las etiquetas de título, junto con otros factores, para tratar de determinar la temática del sitio web. De forma similar, la elección de uno de los resultados de una SERP por parte de los usuarios estará determinada por el grado en que los títulos se ajusten a su intención de búsqueda y por el interés que estos despierten en él.

Existen una serie de aspectos a tener en cuenta a la hora de optimizar las etiquetas de título:

- *Títulos únicos:* Cada página web debe contar con su propio título. Si existen muchas páginas con el mismo título dificultaremos la labor de identificación de la temática de dichas páginas por parte de los motores de búsqueda. Por otra parte, generalmente los títulos repetidos resultan tan genéricos (nombre empresa, nombre sitio web...) que no resultan atractivos para los usuarios, pues no ofrecen información sobre el contenido de la página.
- *Títulos descriptivos:* Además de ser únicos, los títulos deben describir el contenido alojado en sus respectivas páginas web. Esto ayudará tanto a los motores de búsqueda, permitiendo que identifiquen la temática con mayor facilidad, como a los usuarios, que reciben información mucho más precisa sobre el contenido que alberga cada web.
- *Palabras clave primero:* El formato y la localización en la que aparecen los títulos tanto en las páginas de resultados como en las pestañas de los navegadores, hacen que resulte altamente recomendable que las palabras más relevantes de cada una de las páginas aparezcan al comienzo de los títulos. En la Figura 49 podemos ver algunos casos que ejemplifican esta situación:

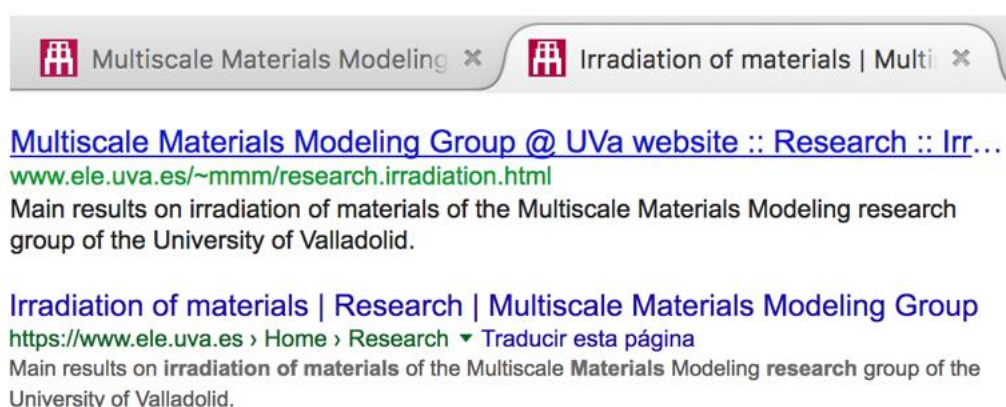


Figura 49. Comparativa de la visualización de la etiqueta de título en sus distintos formatos.

En ambos ejemplos se trata de la misma página, la subsección 'Irradiation of materials', ubicada en la sección 'Research' del sitio web objeto de este proyecto. Como se puede ver, tanto en la pestaña de los navegadores como las páginas de resultados, el hecho de no colocar las principales palabras clave que cubren la temática de la página al comienzo de la etiqueta de título puede provocar que esta quede oculta, como es el caso. En un principio

esto afectará mucho más a los usuarios pues no les permitirá conocer la temática del sitio web.

- *Tamaño:* Se trata de una cuestión de usabilidad y de experiencia de usuario. Los motores de búsqueda destinan un número máximo de caracteres para las etiquetas de título, delimitado por el tamaño reservado para estas en el formato de la página de resultados. Tanto Google como Bing cuentan con un tamaño, de media, de unos 50-60 caracteres. Este valor puede cambiar dependiendo de los caracteres empleados en el título hasta aproximadamente los 35-77 caracteres. Una vez superado el límite de caracteres el buscador trunca el título, de forma que este no se visualiza completamente. Por este motivo cobra mayor importancia el punto anterior, pues si el título es demasiado largo y no se emplean las palabras clave al inicio existe el riesgo de que dichas palabras clave no se muestren a los usuarios. La Figura 50 ejemplifica perfectamente este hecho.

[Multiscale Materials Modeling Group @ UVa website :: Research :: Irr...](http://www.ele.uva.es/~mmm/research.irradiation.html)
www.ele.uva.es/~mmm/research.irradiation.html

Main results on irradiation of materials of the Multiscale Materials Modeling research group of the University of Valladolid.

[Irradiation of materials | Research | Multiscale Materials Modeling Group](https://www.ele.uva.es)

<https://www.ele.uva.es> > Home > Research > Traducir esta página

Main results on irradiation of materials of the Multiscale Materials Modeling research group of the University of Valladolid.

Figura 50. Ejemplo del truncado de la etiqueta de título en las SERPs.

- *Estructura de la página web:* Se recomienda incluir información (como secciones, categorías...) que permita al usuario visualizar la jerarquía del contenido de la página web. Generalmente se sigue el siguiente esquema, empleando el carácter '|' para separar cada una de las partes:

[Nombre de producto] | [Categoría de producto] | [Marca/Empresa]

Ej. *Irradiation of materials | Research | Multiscale Materials Modeling Group*

- *Marca:* Resulta conveniente añadir el nombre de la marca, compañía, ... en los títulos pues nos permite asociar dicha página con la marca, compañía... correspondiente. Especialmente cuando estas son reconocibles, pues su visualización por parte de los usuarios puede determinar la elección de dicha página, mejorando el CTR de la misma, lo que repercutirá en su posicionamiento.

Es recomendable seguir estas pautas, pero aun así Google se reserva el derecho de modificar la etiqueta de título mostrada en las páginas de resultados si considera que esta no se ajusta al contenido de la página web correspondiente.

Los cambios llevados a cabo en las etiquetas de título del sitio web se han realizado bajo las premisas anteriormente expuestas. De hecho, las imágenes comparativas de los ejemplos se corresponden con páginas del propio sitio web, antes y después de los cambios.

El principal problema a subsanar radicaba en el hecho de que el sitio web, originalmente, no emplea las palabras clave al inicio de los títulos, con los problemas que ello conlleva. Y la situación es

especialmente grave pues, tal y como podemos ver en la imagen anterior, debido al gran tamaño del nombre del grupo de investigación, colocado al comienzo del título, la práctica totalidad de las páginas presentaban el mismo título en las páginas de resultados, al nefasto desde el punto de vista de la usabilidad.

Por lo demás, el sitio web original seguía en sus etiquetas de título las pautas anteriormente establecidas, pues estos eran únicos, descriptivos, presentaban la estructura del sitio web, mencionan la 'marca'...

A lo largo del proyecto, y en función de la introducción de otros elementos se han probado diferentes configuraciones en el formato de las etiquetas, en concreto la relativa a la estructura del sitio web. En un principio, tal y como se puede ver en la Figura 51, se incluyeron todos y cada uno de los elementos de la estructura en cada una de las páginas web.

Binary collision approximation | Atomistic modeling techniques...
<https://www.ele.uva.es/~mmm/research/atom.../binary-collision-approximation.html>
Description of the Binary collision approximation, one of the techniques used by the
Multiscale Materials Modeling research group of the University of ...

Figura 51. *Etiqueta de título larga, con todos los elementos de la estructura.*

Esto hacía que en algunas páginas de la sección 'Research', con nombres especialmente largos, el resultado no se ajustará al deseado, pues resultaba difícil apreciar las categorías del sitio web a las que pertenecía el contenido.

Tras la introducción de los Breadcrumbs, tanto en su forma física en las propias páginas web como a través de sus correspondientes datos estructurados –aspecto que veremos más adelante– la situación cambió. Google es capaz de detectar los Breadcrumbs y mostrarlos en las páginas de resultados, tal y como podemos ver en la Figura 52.

Molecular implants | Multiscale Materials Modeling Group | UVa
<https://www.ele.uva.es> > [Home](#) > [Research](#) > [Irradiation of materials](#) ▼ [Traducir esta página](#)
Comparative study of monoatomic and molecular implantations in order to determine the advantages and drawbacks of each approach.

Figura 52. *Etiqueta de título corta. Elementos de la estructura ubicados en el Breadcrumb.*

De esta forma podemos reducir el tamaño de las etiquetas de título sin sacrificar la capacidad de mostrar las diferentes categorías a las que pertenece el contenido de la página y como está jerarquizado el sitio web.

3.2.3.2. Etiqueta de descripción

La etiqueta <meta> description nos permite proporcionar a los motores de búsqueda una descripción del contenido de la página web, que será visible en las páginas de resultados. No se trata de un elemento que mejore directamente el posicionamiento de una página, se trata nuevamente de un elemento estético que nos permite mejorar la experiencia de usuario, pues al aparecer junto con la etiqueta de título en las SERPs supone el primer contacto de los usuarios con la página y puede marcar la diferencia a la hora de obtener visitas. Al igual que ocurriría con las etiquetas de título, una descripción óptima puede mejorar a las métricas de uso de una página, como el Click Through Rate (CTR), y

afectar finalmente a su posición. En la Figura 53 podemos ver un ejemplo de su sintaxis y del aspecto que presenta en las SERPs:

```
<meta name="description" content="We present an extended model for B clustering in crystalline or in preamorphized Si and with validity under conditions below and above the equilibrium solid solubility limit of B in Si." />
```

Kinetics of large B clusters in crystalline and preamorphized silicon
<https://www.ele.uva.es> › Home › Publications ▼
 We present an extended model for B clustering in crystalline or in **preamorphized** Si and with validity under conditions below and above the equilibrium solid ...

Figura 53. Ejemplo etiqueta <meta> description.

A continuación, describiremos una serie de aspectos a tener en cuenta a la hora de optimizar las etiquetas de descripción:

- *Únicas:* Cada página web deberá contar con su propia descripción. Si empleamos la misma descripción genérica para todas nuestras páginas perderemos la ocasión de describir el contenido de la página a los usuarios lo que nos puede llevar a perder visitantes.
- *Descriptivas:* Es importante que se describa correctamente el contenido de la página web, pues se trata de una oportunidad de atraer a los usuarios. Una buena descripción puede marcar la diferencia entre recibir una visita o no recibirla.
- *Emplear palabras clave:* Cuando alguno de los términos de la consulta realizada por el usuario aparece en la descripción, este queda marcado en negrita. Este es un elemento visual muy importante pues, generalmente, llama la atención de los usuarios, y puede aumentar las posibilidades de recibir una visita. En la Figura 54 podemos ver un ejemplo para el término 'photoluminescence':

Insights on the atomistic origin of X and W photoluminescence lines in ...
<https://www.ele.uva.es> › Home › Publications ▼ Traducir esta página
 We have used atomistic simulations to identify and characterize interstitial defect cluster configurations candidate for W and X **photoluminescence** centers in ...

Figura 54. Ejemplo de palabra clave marcada en negrita en la SERP.

- *Tamaño:* Los motores de búsqueda tienen un tamaño medio para la etiqueta description de 160 caracteres, a partir de ese tamaño la descripción queda truncada. En casos puntuales la etiqueta puede llegar a medir más de 200 caracteres, pero no es común.
- *Cuando sean necesarias:* Cuando no hacemos uso de la etiqueta <meta> description, los motores de búsqueda generan la descripción de forma automática, analizando el contenido de la página y tomando aquellas partes que coinciden con la consulta del usuario. Existen casos, especialmente cuando la página web tiene como objetivo el tráfico dentro del 'long-tail', en los que la descripción no resulta necesaria, dado que una descripción estática podría ajustarse peor a las consultas de los usuarios que la que genera el propio motor de búsqueda.

Al igual que ocurre con las etiquetas de título, seguir estas pautas no garantiza que las descripciones se vayan a mostrar tal y como las hemos creado, pues los motores de búsqueda se reservan el derecho de modificarlas si consideran que el contenido de estas no se ajusta a la consulta del usuario.

Las modificaciones llevadas a cabo en las etiquetas <meta> description del sitio web objeto de este proyecto se han realizado bajo las premisas anteriormente descritas. En origen el sitio web seguía las premisas correctamente, con descripciones únicas, descriptivas, en las que se emplean las palabras clave correspondientes, con el tamaño correcto...

Por lo tanto, el trabajo se ha centrado en crear las descripciones de las nuevas páginas del sitio web, pertenecientes a las secciones 'Research', 'Projects' y 'Publications', especialmente la última sección. Dado que tienen como objetivo palabras clave pertenecientes al 'long-tail', se han llevado a cabo diferentes pruebas relativas al tamaño de la etiqueta, a la necesidad de su incorporación... Finalmente se optó por mostrar el abstract correspondiente a cada una de las publicaciones.

3.2.3.3. Etiquetas de encabezado

Las etiquetas de encabezado HTML(<h1> - <h6>) están diseñadas para indicar la jerarquía existente en los títulos de un documento y, de esta forma, dotar al contenido de una estructura que permita identificar sus principales secciones tanto a los motores de búsqueda como por los usuarios. Los motores de búsqueda emplean las etiquetas de cabecera para determinar la temática que cubre el contenido de una página web. Si las palabras clave para las que deseamos posicionar una página se encuentran en las etiquetas de cabecera la probabilidad de aparecer en los resultados de búsqueda para las consultas que empleen dichas palabras clave aumentará. No por ello se debe hacer un uso irresponsable de estas etiquetas colocando demasiado contenido, pues entonces perderán su utilidad.

Las etiquetas de encabezado presentan seis niveles, que van de <h1> a <h6>. La etiqueta <h1> suele considerarse como el título de la página, la <h2> se reserva para los subtítulos dentro de un documento, la <h3> para los subtítulos de tercer nivel, y así sucesivamente, reduciendo su relevancia a nivel que profundizamos en la jerarquía hasta llegar a la etiqueta <h6>. Cuando una página cuenta con varias etiquetas de encabezado los motores tendrán en mayor consideración las etiquetas de mayor nivel en la jerarquía frente a las de menor nivel.

Es recomendable hacer uso de una única etiqueta <h1> y mantener, siempre que sea posible, un orden a la hora de emplear las etiquetas de forma que las etiquetas de mayor nivel precedan a la de menor nivel y no al revés, a fin de mantener un estructura coherente y fácil de interpretar.

En el caso que nos ocupa, el principal error a subsanar es que el sitio web original empleaba la etiqueta <h1> en el logo del sitio web, por lo que todas las páginas tenían exactamente la misma etiqueta <h1>. Esto soluciona modificando todas las páginas y dotando a las etiquetas <h1> con el contenido correcto.

En cuanto al uso del resto de etiquetas, este está sujeto a las necesidades que presente el contenido de cada una de las páginas, por ello emplearemos distintas configuraciones de etiquetas. Veamos algunos ejemplos, en el caso de las nuevas páginas de la sección 'Research', que cuentan con un apartado de 'Publicaciones Relacionadas', hacemos uso de las etiquetas <h1>, <h2> y <h3>, como podemos ver en la Figura 55.



Figura 55. Etiquetas de encabezado empleadas en las nuevas páginas de la sección 'Research'.

En el caso de las páginas web pertenecientes a cada una de las publicaciones, emplearemos otra configuración. En concreto haremos uso de las etiquetas <h1>, <h2>, <h3> y <h4>, tal y como se puede ver en la Figura 56.

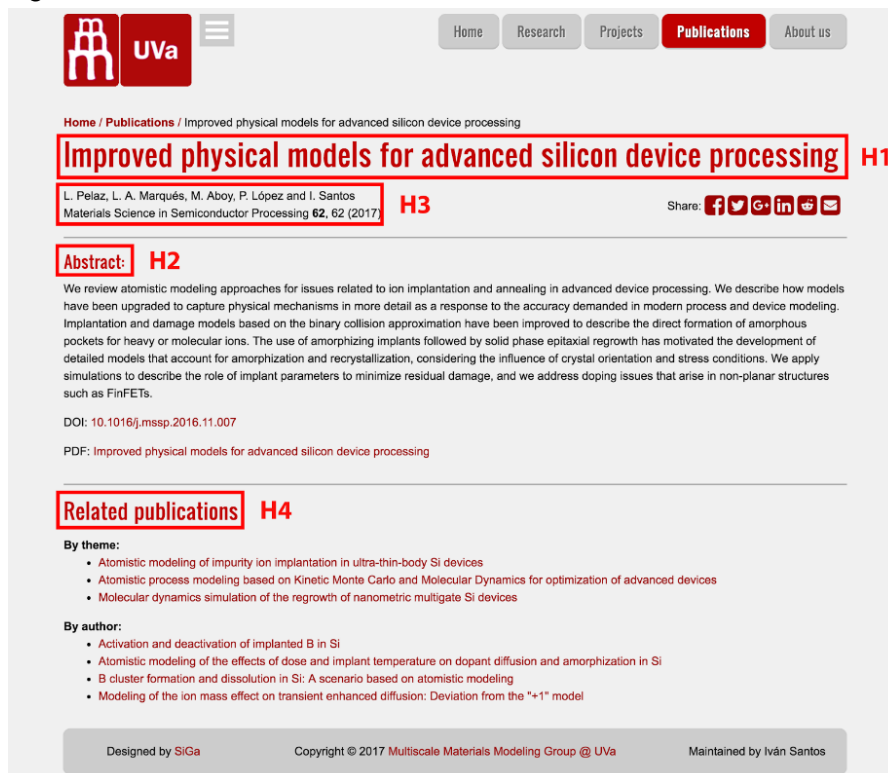


Figura 56. Etiquetas de encabezado empleadas en las páginas de las publicaciones.

3.2.3.4. Nombre de archivo y atributo *alt* de las imágenes

Incorporar imágenes suele enriquecer la experiencia de usuario, pero los motores de búsqueda no tienen la capacidad de ver las imágenes tal y como lo haría un humano. Existen dos elementos que permiten a los motores de búsqueda establecer el contexto de una imagen y de esta forma interpretar su contenido:

- **Nombre de archivo:** Los motores de búsqueda analizan el nombre de archivo de las imágenes para tratar de obtener alguna pista que les permita conocer su contenido. Por tanto, no debemos desaprovechar la oportunidad de incluir keywords en los nombres de archivo de las imágenes, en lugar de emplear términos irreconocibles como en el siguiente ejemplo:

<https://www.ele.uva.es/~mmm/img/img2348901.png>
<https://www.ele.uva.es/~mmm/img/res-irr-molecular-implants.png>

- **Atributo *alt*:** Las etiquetas de imagen de HTML () permiten hacer uso del atributo *alt*. En el podemos proporcionar más información sobre el contenido de la imagen. A continuación, tenemos un ejemplo de la sintaxis que presenta:

``

El uso de estos elementos, además de mejorar la forma en la que los motores de búsqueda interpretan las imágenes, nos permite reforzar el uso de las palabras clave hacia las que están enfocadas dichas páginas y para las que deseamos lograr una mejor posición.

3.2.3.5. Anchor text

Cuando hablamos de anchor text o “texto de anclaje” nos referimos al texto visible de un enlace. Los motores de búsqueda consideran los anchor text como la forma en la que se describe el contenido enlazado. Por tanto, si una página recibe numerosos enlaces con el anchor text “libros” es un claro indicativo de que el contenido de dicha página trata sobre libros.

Existen dos tipos de enlaces, los internos, que sirven para navegar dentro de un mismo sitio web, y que determinan la estructura del mismo. Y, por otro lado, tenemos los enlaces externos, aquellos recibidos desde otros sitios web y sobre los que no tenemos control. Haremos un breve inciso sobre los enlaces externos más adelante, en la Sección 3.3.1.

En nuestro caso nos centraremos en los textos de enlace internos. Estos son algunos de aspectos a tener en cuenta:

- **Sucintos:** No existe un tamaño estándar, pero deberían ser tan cortos como sea posible.
- **Descriptivos:** Deben describir el contenido de la página a la que enlaza. En el caso de los enlaces internos, los anchor text vienen marcados por la propia estructura del sitio web y las palabras clave sobre las que está construida.
- **Variados:** En este caso no es necesario que sean únicos, pero sí que tengan cierta variedad. Un número demasiado elevado de anchor text iguales puede ser considerado por los motores de búsqueda como spam, tanto para los enlaces internos como los externos. Por tanto, deberemos evitar en la medida de lo posible los títulos genéricos.

- *Palabras clave:* Hay que hacer uso coherente de las palabras clave empleadas en los anchor text. Penguin persigue la sobre optimización del anchor text y lo considera como una técnica de spam, tanto para los enlaces internos como los externos.

El general el sitio web original seguía estas directrices perfectamente, salvo por una cuestión, de la que ya dimos cuenta en la auditoría previa. A la hora de enlazar los artículos en versión en versión PDF emplea exactamente el mismo anchor text ([PDF]) y por tanto no respeta los aspectos vistos anteriormente. En la Figura 57 tenemos un ejemplo de esa situación:

Insights on the atomistic origin of X and W photoluminescence lines in c-Si from ab initio simulations
I. Santos, M. Aboy, P. López, L. A. Marqués, and L. Pelaz
[Journal of Physics D: Applied Physics 49, 075109 \(2016\), \[PDF\]](#)

Molecular dynamics simulation of the early stages of self-interstitial clustering in silicon
L. A. Marqués, I. Santos, L. Pelaz, P. López, and M. Aboy
[Materials Science in Semiconductor Processing 42, 235 \(2016\), \[PDF\]](#)

Atomistic modeling of ion implantation technologies in silicon
L. A. Marqués, I. Santos, L. Pelaz, P. López, and M. Aboy
[Nuclear Instruments and Methods in Physics Research B 352, 148 \(2015\), \[PDF\]](#)

Figura 57. *Uso incorrecto del anchor text.*

La solución pasa por emplear los títulos de las publicaciones para enlazar su respectiva versión PDF. Dado que con el proceso de reestructuración de la sección ‘Publications’ nos ha llevado a crear una nueva página para cada una de las publicaciones la situación es algo diferente. En la Figura 58 podemos ver que la página principal de la sección ‘Publications’ ahora enlaza cada publicación con su respectiva página web.

Insights on the atomistic origin of X and W photoluminescence lines in c-Si from ab initio simulations
I. Santos, M. Aboy, P. López, L. A. Marqués and L. Pelaz.
[Journal of Physics D: Applied Physics 49, 075109 \(2016\)](#)

Molecular dynamics simulation of the early stages of self-interstitial clustering in silicon
L. A. Marqués, I. Santos, L. Pelaz, P. López and M. Aboy.
[Materials Science in Semiconductor Processing 42, 235 \(2016\)](#)

Atomistic modeling of ion implantation technologies in silicon
L. A. Marqués, I. Santos, L. Pelaz, P. López and M. Aboy.
[Nuclear Instruments and Methods in Physics Research B 352, 148 \(2015\)](#)

Figura 58. *Uso correcto del anchor text para enlazar las páginas de las publicaciones.*

Y es dentro de cada una de estas páginas donde enlazamos con la versión en PDF de la respectiva publicación, junto con el enlace a la revista científica donde fue publicado. Como se puede ver en Figura 59, el anchor text se corresponde con el título de la publicación en lugar de emplear uno genérico como [PDF]:

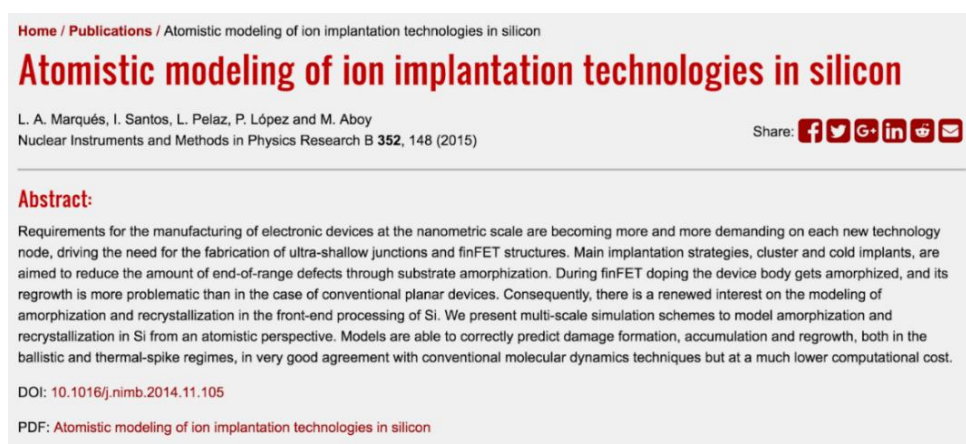


Figura 59. Uso correcto del anchor text para enlazar los archivos PDF.

3.2.3.6. Datos estructurados

Cuando hablamos de datos estructurados nos referimos a tipos de datos con un alto nivel de organización, de forma similar a la información contenida en una base de datos relacional. Cuando la información está altamente estructurada y es predecible, los motores de búsqueda pueden organizarla con mayor facilidad y de formas más creativas [\[89\]](#).

Con el fin de dar a la información relevante de nuestros sitios web una estructura y un formato predecible, emplearemos un lenguaje de marcado de datos estructurados. Emplearemos el vocabulario que proporciona Schema.org, una iniciativa lanzada en 2011 por Google, Bing, Yahoo! y Yandex, y que tiene como objetivo crear y dar soporte a un conjunto de esquemas que permitan el marcado de datos estructurados en páginas web.

El objetivo es emplear el vocabulario y los esquemas de Schema.org, junto con los diferentes formatos soportados, para marcar el contenido de un sitio web. Este marcado es reconocido por los robots de los motores de búsqueda y permite a los motores de búsqueda situar la información obtenida en un contexto y otorgarle un significado que de otra forma no obtendría.

El marcado de datos estructurados describe entidades de la Web, junto con sus propiedades. Si como en nuestro caso un sitio web contiene artículos, podemos emplear el marcado para describir las propiedades de artículo, es decir, su título, autores, descripción, fecha de publicación...

Una vez que el contenido del sitio web ha sido marcado este puede aparecer en dos de las categorías dentro de las funcionalidades de Google Search:

- *Resultados enriquecidos (Rich results)*: cuando se emplean sobre recetas, noticias o videos estos pueden aparecer en Rich Cards, como un elemento único o formando parte de una lista. Existen otros tipos de datos estructurados que nos permiten mejorar la apariencia del sitio web en los resultados de búsqueda, como pueden ser los Breadcrumbs o las valoraciones. En la Figura 60 tenemos un ejemplo de esto último.



Figura 60. Ejemplo de resultado enriquecido con el marcado de la valoración.

- **Knowledge Graph cards:** Si un sitio web es la autoridad para cierto contenido, Google puede tratar los datos estructurados del sitio web como hechos e incluirlos en el Knowledge Graph y, de esta forma este contenido puede aparecer en respuestas destacadas. Las tarjetas del Knowledge Graph aparecen para proporcionar datos autorizados sobre organizaciones, eventos, personas, cosas... tal y como tenemos en la Figura 61.

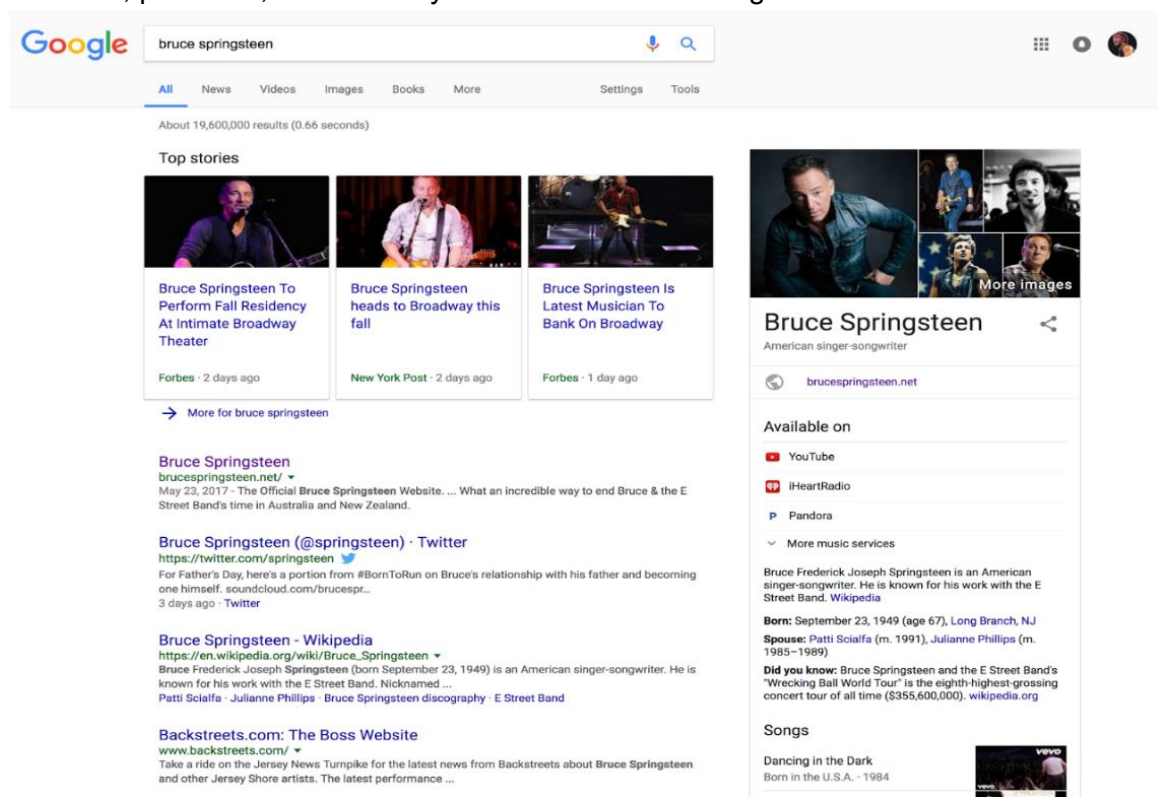


Figura 61. Ejemplo de Knowledge Graph.

3.2.3.6.1. Lenguajes de marcado

Existen varios lenguajes para el marcado de datos estructurados. A continuación, describiremos los principales:

- **Microdata**

Es una especificación HTML empleada para incluir datos estructurados dentro del código HTML. Para ello emplea atributos de las etiquetas HTML para determinar las propiedades que se deseen incluir como datos estructurados. Soporta todos los tipos de datos.

- *RDFa*

Se trata de una extensión HTML5 que permite la vinculación de datos mediante la introducción de atributos en las etiquetas HTML, correspondiendo estas con el contenido visible por el usuario, de forma similar a como ocurría con el formato Microdata. Soporta todos los tipos de datos.

- *JSON-LD (JavaScript Object Notation for Linked Data)*

Se trata de una notación JavaScript separada del cuerpo del archivo HTML. El marcado se sitúa entre etiquetas de script (<script>...</script>) en la cabecera del archivo HTML. No es necesario que el marcado se entrelaza con el texto visible por el usuario, lo que hace que los datos anidados como pueden ser una dirección postal sean más fáciles de expresar. Además, Google tiene la capacidad de leer datos en formato JSON-LD inyectados dinámicamente en una página mediante código JavaScript o con un widget. Soporta todos los tipos de datos exceptos los Breadcrumbs.

Google recomienda el uso de JSON-LD, que está en proceso de soportar todos los tipos de datos. Además, como hemos visto en los ejemplos sobre estas líneas, se trata de la forma más simple de incluir datos estructurados, pues no requiere modificar el código HTML existente.

3.2.3.6.2. Tipos de datos estructurados

Existen múltiples tipos de datos estructurados, en esta sección únicamente describiremos los empleados en el sitio web. Dividiremos los tipos en dos apartados, por un lado, los datos estructurados relativos información del sitio web y a su estructura, y por otro los relativos al marcado del contenido.

3.2.3.6.2.1. Información y estructura del sitio web

Este apartado contiene los datos estructurados empleados para marcar los datos relativos a la estructura e información del sitio web.

- **Breadcrumbs**

Los breadcrumbs, o migas de pan, nos permiten indicar a los usuarios el punto en el que se encuentran dentro de la jerarquía del sitio web, facilitando su navegación. Cuando marcamos este tipo de datos estructurados los motores de búsqueda lo detectarán y mostrarán el breadcrumb en las SERPs, tal y como se puede ver en la Figura 62.

Molecular implants | Multiscale Materials Modeling Group | UVA
<https://www.ele.uva.es> > Home > Research > Irradiation of materials ▼
Comparative study of monoatomic and **molecular** implantations in order to determine the advantages and drawbacks of each approach.

Figura 62. Ejemplo del marcado de los breadcrumbs.

Para el marcado de los Breadcrumbs se empleará la entidad *BreadcrumbList*. El conjunto de propiedades de esta entidad, junto con sus tipos y la pertinente descripción de cada una de ellas, se encuentra en su correspondiente página en la web de Schema.org [90].

Dado que el formato JSON-LD, el formato de marcado recomendado por Google, de momento no da soporte al marcado de Breadcrumbs emplearemos el formato Microdata. En la Figura 63 tenemos

un ejemplo completo de su sintaxis, correspondiente con el marcado de uno de los breadcrumbs del sitio web.

```
<ol class="nav breadcrumb" itemscope itemtype="https://schema.org/BreadcrumbList">
  <li itemprop="itemListElement" itemscope itemtype="https://schema.org/ListItem">
    <a itemscope itemtype="http://schema.org/Thing" itemprop="item" href="https://www.ele.uva.es/~mmm/">
      <span itemprop="name">Home</span></a>
      <meta itemprop="position" content="1" />
    </li>
  <li itemprop="itemListElement" itemscope itemtype="https://schema.org/ListItem">
    <a itemscope itemtype="http://schema.org/Thing" itemprop="item" href="https://www.ele.uva.es/~mmm/research/">
      <span itemprop="name">Research</span></a>
      <meta itemprop="position" content="2" />
    </li>
  <li itemprop="itemListElement" itemscope itemtype="https://schema.org/ListItem">
    <a itemscope itemtype="http://schema.org/Thing" itemprop="item" href="https://www.ele.uva.es/~mmm/research/atomistic-modeling-techniques/">
      <span itemprop="name">Atomistic modeling techniques</span></a>
      <meta itemprop="position" content="3" />
    </li>
  <li>Tight-binding</li>
</ol>
```

Figura 63. Ejemplo de la sintaxis del marcado de un breadcrumb. (Microdata)

Como podemos ver las propiedades van incluidas en las etiquetas HTML mediante la utilización de atributos. Por tanto, el código se incluirá allí donde esté situado el breadcrumb, al contrario que ocurre cuando empleamos JSON-LD, que es independiente.

- Información del sitio web

Este tipo de dato estructurado se emplea para ofrecer información sobre el sitio web a los motores de búsqueda. Google toma el nombre proporcionado y en ocasiones lo muestra en las SERPs sustituyendo al dominio, tal y como se puede ver en Figura 64.



Figura 64. Ejemplo del uso en las SERPs del marcado de la info. de sitio web.

Para el marcado de la información de un sitio web se empleará la entidad *WebSite*. El conjunto de propiedades de esta entidad, junto con sus tipos y la pertinente descripción de cada una de ellas, se encuentra en su correspondiente página en la web de Schema.org [91].

En este caso emplearemos el formato JSON-LD, recomendado por Google. En la Figura 65 podemos observar un ejemplo de su sintaxis, correspondiente con el marcado de la página principal. Incluimos este código únicamente en dicha página.

```
<script type="application/ld+json">
  {
    "@context": "http://schema.org",
    "@type": "WebSite",
    "name": "Multiscale Materials Modeling Group | UVa",
    "alternateName": "MMMG-UVa",
    "url": "https://www.ele.uva.es/~mmm/"
  }
</script>
```

Figura 65. *Ejemplo de la sintaxis del marcado de la info. de sitio web. (JSON-LD)*

- Marcado de la información del negocio

Este tipo de esquema nos permite mejorar la apariencia y visibilidad de la compañía. Para ello marcaremos datos como información de contacto, dirección postal, logos, perfiles en redes sociales... de forma que estos aparezcan en los resultados de búsqueda, en las tarjetas del Knowledge Graph y en Google Maps.

Existe otra forma de proporcionar estos datos a los motores de búsqueda, a través de Google My Business. Pero dada la naturaleza de la empresa que nos ocupa en este proyecto, es decir, un grupo de investigación perteneciente a la Universidad de Valladolid, no parece que Google My Business se ajuste a nuestras necesidades.

A la hora de marcar la información de negocio emplearemos la entidad 'Organization', o cualquiera de sus subtipos, como 'GovernmentOrganization', 'MedicalOrganization', 'EducationalOrganization'... Seleccionaremos la que mejor se ajuste a nuestras necesidades, en nuestro caso 'EducationalOrganization'. El conjunto de propiedades de esta entidad, junto con sus tipos y la pertinente descripción de cada una de ellas, se encuentra en su correspondiente página en la web de Schema.org [\[92\]](#). En nuestro caso emplearemos las propiedades que mejor se ajusten a la información de la que disponemos.

Para el marcado de los datos de la entidad 'EducationalOrganization' se ha empleado el lenguaje de marcado JSON-LD. En la Figura 66 de la sintaxis, correspondiente con el marcado de la información de nuestro sitio web.

```

<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "EducationalOrganization",
  "name": "Multiscale Materials Modeling Group - UVa",
  "url": "https://www.ele.uva.es/~mmm/",
  "logo": "https://www.ele.uva.es/~mmm/img/logo_mmm.png",
  "address": {
    "@type": "PostalAddress",
    "addressCountry": "Spain",
    "addressLocality": "Valladolid",
    "postalCode": "47011",
    "streetAddress": "Dpt. Electrónica, E.T.S.I. Telecomunicación, Campus Miguel
Delibes, Paseo Belén 15"
  },
  "contactPoint": {
    "@type": "ContactPoint",
    "telephone": "(+34) 983 423 683 (ext. 5512)",
    "contactType": "customer service",
    "email": "ivasan@tel.uva.es"
  },
  "email": "ivasan@tel.uva.es",
  "location": {
    "@type": "Place",
    "hasMap": "https://www.google.com/maps/d/embed?mid=z-Gd4S2Sxjho.kWidKCzFQ5z4"
  },
  "parentOrganization": {
    "@type": [ "CollegeOrUniversity",
      "EducationalOrganization",
      "Organization"
    ],
    "name": "Universidad de Valladolid",
    "url": "http://www.uva.es"
  }
}
</script>

```

Figura 66. Ejemplo de la sintaxis del marcado de la info. del negocio (JSON-LD)

Este código se colocará en la cabecera de la página principal del sitio web de la organización y en la página de contacto.

3.2.3.6.2.2. Contenido del sitio web

En este apartado procederemos a marcar los datos estructurados relativos al contenido de las páginas web. Cuando empleamos el marcado de datos estructurados para marcar el contenido de un sitio web permitimos que Google entienda mejor el contexto de dicho contenido de forma que este se muestre apropiadamente en las SERPs. Para ello marcaremos las propiedades del contenido (título, fecha, autor, director, lugar...) y habilitaremos acciones (reserva restaurante, visualización de contenido, pedidos de productos) cuando proceda. Al marcar estos contenidos permitimos que aparezcan en los resultados enriquecidos, a través de Rich Cards, o en las Knowledge Graph Cards, lo que proporciona mayor visibilidad al contenido.

- Tipos de contenido

Existe una variedad de tipos de contenido que son susceptibles de aparecer en los resultados enriquecidos y en las Knowledge Graph Cards. Buena parte de ellos se engloban bajo el tipo *CreativeWork* de Schema.org [93], que se aplica a contenido producido para ser leído, escuchado, visualizado o cualquier otro tipo de consumo, como pueden ser videos, libros, noticias... Por otro lado, tenemos los tipos de contenido relativos al comercio, como pueden ser los negocios locales, sus productos y eventos relacionados con los mismos.

En la Tabla 6 podemos ver los tipos de contenido que Google destaca, junto con las características que se habilitan una vez que el contenido ha sido marcado. Algunos de ellos requieren una solicitud explícita.

Tabla 6. *Tipos de contenido destacados por Google.*

Tipo de contenido	Características disponibles	Notas
Artículos	Rich cards	
Negocios locales	Acciones relativas al negocio	Requiere solicitud
Libros	Acciones relativas a los libros	
Cursos	Rich cards	
Música	Acciones relativas a la música	Requiere solicitud
Recetas	Rich cards. Listas de recetas.	
Críticas / Review	Critic review cards	Requiere solicitud
Películas & TV	Acciones relativas al visionado	Requiere solicitud
Videos	Rich cards	

Todos ellos pueden ser marcados empleando los lenguajes de marcado vistos anteriormente. Además, Google Search Console cuenta con una herramienta que nos permite marcar de forma automática el contenido de las páginas web pero que, tal y como se puede ver en la Figura 67, limita los tipos de información a marcar a tipos similares a los mencionados anteriormente.

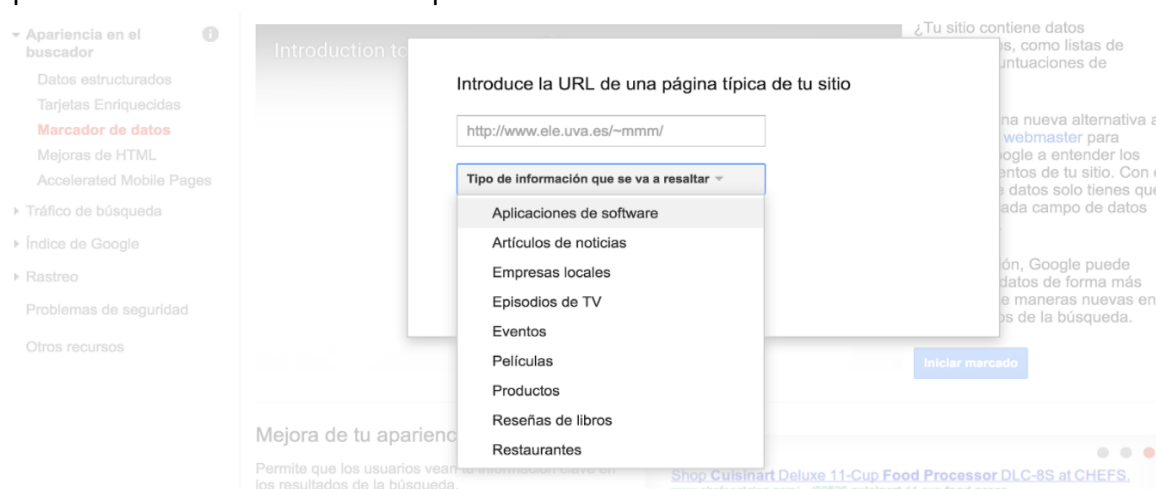


Figura 67. *Herramienta para el marcado automático de datos estructurados. Google Search Console.*

En principio ninguno de estos tipos de contenido, sobre los que Google pone el foco, se ajusta al contenido con el que contamos en el sitio web del que es objeto este proyecto. El que más se le acerca son los artículos de noticias, pero su carácter nada tiene que ver con el de los artículos académicos. En una noticia prima la inmediatez y su vigencia es mucho menor que la de un artículo académico, que puede ser una referencia durante décadas. Por ello el contenido a marcar es muy diferente.

Para el marcado de los artículos de noticias Google emplea el esquema “NewsArticle” de Schema.org [94]. Se trata de una especificación del esquema “Article” [95]. Nosotros emplearemos otra de sus especificaciones, el esquema “ScholarlyArticle”, especialmente indicado para artículos académicos.

- ScholarlyArticle

Nuestro objetivo a la hora de emplear este tipo de esquema es aumentar la visibilidad de las publicaciones del grupo de investigación de forma que Google pueda reconocer e indexar su contenido con mayor facilidad y celeridad.

Antes de comenzar con el marcado, en la Figura 68 podemos ver el aspecto de la página HTML de una de las publicaciones del grupo de investigación. En ella podemos ver el contenido que debe ser marcado.

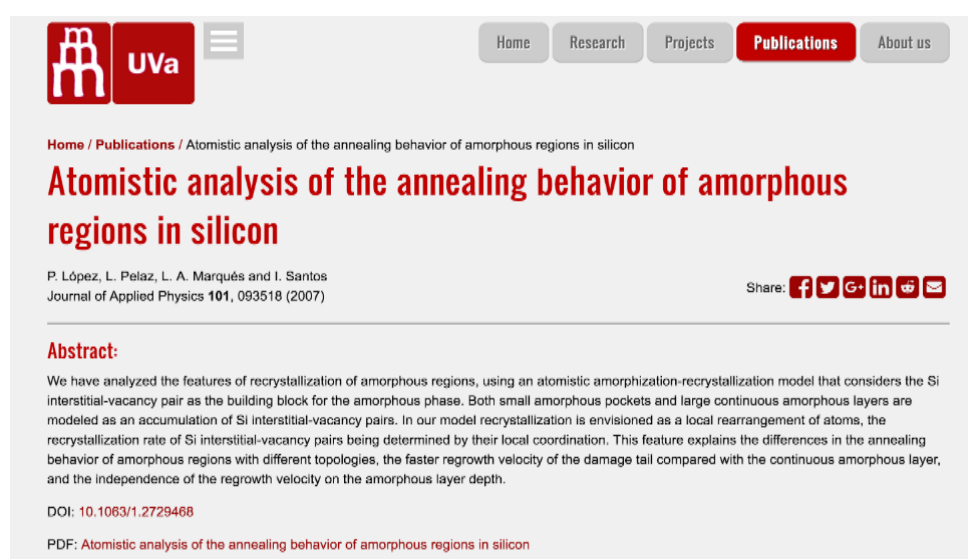


Figura 68. Página web correspondiente a una de las publicaciones.

Como ya hemos comentado, para el marcado de estos datos emplearemos el esquema correspondiente con la entidad ‘ScholarlyArticle’. El conjunto de propiedades de esta entidad, junto con sus tipos y la pertinente descripción de cada una de ellas, se encuentra en su correspondiente página en la web de Schema.org [96]. En nuestro caso emplearemos las propiedades que mejor se ajusten a nuestro contenido.

Para el marcado de las entidades ‘ScholarlyArticle’ hemos empleado dos de los tipos de lenguaje de marcado mencionados anteriormente, Microdata y JSON-LD, con el fin de determinar si existen diferencias a la hora de ser reconocidos por los motores de búsqueda. En las siguientes imágenes podemos ver un ejemplo del aspecto que presentan cada uno de los tipos de lenguaje de marcado sobre el contenido del mismo artículo. Como se puede observar a simple vista la diferencia en la sintaxis empleada es notable.

En primer lugar, tenemos el marcado en JSON-LD, bastaría con colocar el código de la Figura 69 en la cabecera de la página web del artículo correspondiente para marcar su contenido.

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@graph": [
    {
      "@type": "ScholarlyArticle",
      "isPartOf": {
        "@id": "#issue",
        "@type": "PublicationIssue",
        "issueNumber": "9",
        "datePublished": "2007",
        "isPartOf": {
          "@type": [
            "PublicationVolume",
            "Periodical"
          ],
          "name": "Journal of Applied Physics",
          "volumeNumber": "101"
        }
      },
      "name": "Atomistic analysis of the annealing behavior of amorphous regions in silicon",
      "headline": "Atomistic analysis of the annealing behavior of amorphous regions in silicon",
      "datePublished": "2007",
      "pageStart": "093518",
      "description": "We have analyzed the features of recrystallization of amorphous regions, using an atomistic amorphization-
recrystallization model that considers the Si interstitial-vacancy pair as the building block for the amorphous phase. Both small
amorphous pockets and large continuous amorphous layers are modeled as an accumulation of Si interstitial-vacancy pairs. In our model
recrystallization is envisioned as a local rearrangement of atoms, the recrystallization rate of Si interstitial-vacancy pairs being
determined by their local coordination. This feature explains the differences in the annealing behavior of amorphous regions with
different topologies, the faster regrowth velocity of the damage tail compared with the continuous amorphous layer, and the independence
of the regrowth velocity on the amorphous layer depth.",
      "author": [{
        "@type": "Person",
        "name": "Pedro López"
      },
      {
        "@type": "Person",
        "name": "Lourdes Pelaz"
      },
      {
        "@type": "Person",
        "name": "Luis A. Marqués"
      },
      {
        "@type": "Person",
        "name": "Iván Santos"
      }
    ],
      "sameAs": "https://dx.doi.org/10.1063/1.2729468",
      "image": {
        "@type": "ImageObject",
        "url": "https://www.ele.uva.es/~mmm/img/logo_mmm.png",
        "width": "78",
        "height": "100"
      }
    }
  ]
}
</script>
```

Figura 69. Ejemplo de la sintaxis de ScholarlyArticle (JSON-LD)

En cambio, en el código de la Figura 70, en el que empleamos Microdata, tenemos que los atributos de las propiedades con los que marcamos el contenido se incluyen dentro de las etiquetas HTML y por tanto es necesario modificar el código original de la página web objeto del marcado.

```

<div itemscope itemtype="https://schema.org/ScholarlyArticle">
  <h1 itemprop="headline name">Atomistic modeling of defect evolution in Si for amorphizing and subamorphizing implants</h1>

  <strong>Authors:</strong>
  <span itemprop="author" itemscope itemtype="https://schema.org/Person">
    <span itemprop="name">P. López</span>,
  </span>
  <span itemprop="author" itemscope itemtype="https://schema.org/Person">
    <span itemprop="name">L. Pelaz</span>,
  </span>
  <span itemprop="author" itemscope itemtype="https://schema.org/Person">
    <span itemprop="name">L. A. Marqués</span>,
  </span>
  <span itemprop="author" itemscope itemtype="https://schema.org/Person">
    <span itemprop="name">I. Santos</span>,
  </span>
  <span itemprop="author" itemscope itemtype="https://schema.org/Person">
    <span itemprop="name">M. Aboy</span>
  </span>
  and
  <span itemprop="author" itemscope itemtype="https://schema.org/Person">
    <span itemprop="name">J. Barbolla</span>
  </span><br/>

  <strong>Publication:</strong>
  <span itemscope itemtype="https://schema.org/Periodical" itemid="#periodical">
    <span itemprop="name">Materials Science and Engineering: B</span>,
  </span>
  Vol.<span itemprop="isPartOf" itemscope itemtype="https://schema.org/PublicationVolume">
    <link itemprop="isPartOf" href="#periodical"/>
    <span itemprop="volumeNumber">114</span>-<span itemprop="volumeNumber">115</span>,
  </span>
  p.<span itemprop="pageStart">82</span> (<span itemprop="datePublished">2004</span>)

  <div itemprop="image" itemscope itemtype="https://schema.org/ImageObject">
    <meta itemprop="url" content="https://www.ele.uva.es/~mmm/img/logo_mmm.png">
    <meta itemprop="width" content="78">
    <meta itemprop="height" content="100">
  </div>

  <h2>Abstract:</h2>
  <p itemprop="description">Solid phase epitaxial regrowth of pre-amorphizing implants has received significant attention as a method to achieve high dopant activation with minimal diffusion at low implant temperatures and suppress channelling. Therefore, a good understanding of the amorphization and regrowth mechanisms is required in process simulators. We present an atomistic amorphization and recrystallization model that uses the interstitial-vacancy (I-V) pair as a building block to describe the amorphous phase. I-V pairs are locally characterized by the number of neighbouring I-V pairs. This feature captures the damage generation and the dynamical annealing during ion implantation, and also explains the annealing behaviour of amorphous layers and amorphous pockets.</p>
  <p>DOI: <a itemprop="sameAs" href="https://dx.doi.org/10.1016/j.mseb.2004.07.040">10.1016/j.mseb.2004.07.040</a></p>
</div>

```

Figura 70. Ejemplo de la sintaxis de ScholarlyArticle (Microdata)

3.2.3.6.3. Herramientas y Testing

Google pone a disposición de los usuarios la Herramienta de prueba de datos estructurados (Structured Data Testing Tool) [97]. Esta herramienta nos permite determinar si el marcado se ha realizado correctamente. Para ello basta con incluir la URL de la página web marcada, o bien el código HTML de dicha página.

Como resultado se mostrarán aquellas entidades que Google reconozca, así como las propiedades correspondientes al contenido marcado. De existir algún error en la sintaxis, o bien la falta de alguna propiedad de carácter obligatorio, la herramienta lo notificará a la hora de mostrar los resultados del test.

En la Figura 71 podemos ver un ejemplo del correcto funcionamiento de la herramienta. En este caso se trata del marcado de un sitio web, y cómo se puede observar la herramienta reconoce la entidad WebSite junto con las propiedades de la misma.

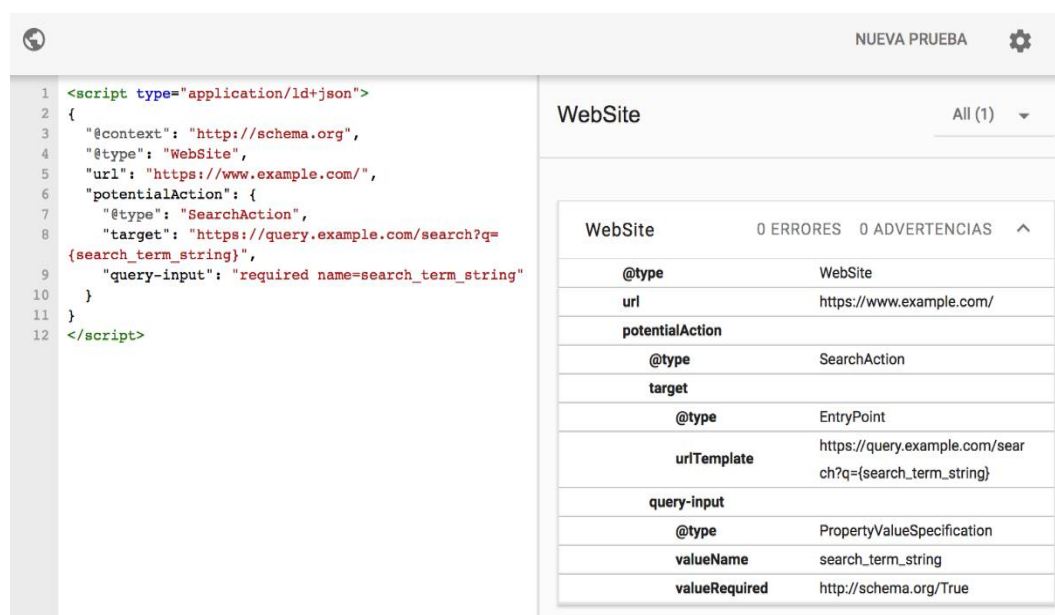


Figura 71. Ejemplo de uso de la herramienta Structured Data Testing Tool.

Todos los tipos de marcado de datos estructurados que hemos visto anteriormente, y que se han incluido en las páginas web del proyecto, han sido probados en esta herramienta para evitar posibles errores de sintaxis y para comprobar que las propiedades incluidas en cada entidad son las correctas.

Por otro lado, tenemos que las herramientas de administración de Google, Google Search Control, cuenta con un apartado en el que muestra la detección de los datos estructurados por parte del motor de búsqueda. De esta forma podremos seguir la evolución de los mismos.

3.2.3.6.4. Resultados del marcado de datos estructurados

Como hemos visto hemos empleado distintos tipos de datos estructurados y lenguaje para el marcado de los datos con el fin de determinar si existen diferencias palpables a la hora de su detección por parte de los motores de búsqueda, en especial Google Search Console (GSC).

Las entidades *WebSite* y *EducationalOrganization* han sido detectadas sin problemas a lo largo de todo el proyecto de forma constante. Aunque por desgracia no parece que hayan surgido ningún efecto en las SERPs.

En el caso de las entidades *BreadcrumbList*, la detección se ha limitado a algo más de la mitad de todas las páginas donde se ha aplicado, pero no de forma constante a lo largo de todo el proyecto. Frente a esta falta de constancia en la detección por parte de GSC, nos encontramos con que para todas las páginas del sitio web que emplean Breadcrumbs, estos aparecen en las SERPs, independientemente de si fueron detectados o no. Esto nos hace dudar de la fiabilidad de la herramienta de detección.

En el caso de las entidades *ScholarlyArticle*, al emplear dos tipos de lenguaje de marcado a lo largo del proyecto, se ha seguido especialmente la evolución en la detección, por parte de GSC, del marcado de dichas entidades, correspondientes a cada una de las publicaciones. Como resultado hemos obtenido que alrededor de la mitad de las publicaciones fueron detectadas en algún momento por GSC,

pero no con la constancia ni con permanencia deseable. Algunas de ellas han ido apareciendo y desapareciendo de la herramienta de detección a lo largo del proyecto. Además, el número de publicaciones detectadas que fueron marcadas con Microdata es similar al de aquellas que fueron marcadas con JSON-LD. Por tanto, los resultados obtenidos no parecen indicar que el lenguaje de marcado empleado sea significativo a la hora de la detección de las entidades *ScholarlyArticle*.

3.2.3.7. Etiquetas Open Graph y Twitter cards

Se trata de un conjunto de etiquetas que nos permiten estandarizar los elementos de un sitio web con el fin de que estos sean interpretados correctamente por las redes sociales como Facebook (Open Graph) y Twitter (Twitter Cards).

De esta forma podemos proporcionar la información (título, descripción, foto, autor, ...) que queremos que se visualice cuando alguien comparte nuestro contenido en las redes sociales, en lugar de dejar dicha tarea a la capacidad de interpretación de las propias redes sociales. Se trata nuevamente de una cuestión estética que puede influir al posicionamiento de forma indirecta, es decir, atrayendo nuevas visitas...

Tanto las etiquetas de Open Graph como las de Twitter Cards cuentan con propiedades similares. A continuación, describiremos dichas propiedades junto con su sintaxis:

- Open Graph

Cuando se colocan estas etiquetas en una página web convertimos dicha página en un objeto del Open Graph. Cuando se emplea en Facebook permite otorgarle a cualquier página web la misma funcionalidad que cualquier objeto de Facebook. Emplearemos las propiedades que mejor se ajustan a nuestro contenido, algunas de ellas obligatorias:

- *og:type*. Determina el tipo del objeto y en función del mismo puede que se requieran más propiedades. (Obligatoria)
- *og:title*. Título del objeto. (Obligatoria)
- *og:url*. La URL canónica, utilizada para identificar al objeto en el Open Graph. (Obligatoria)
- *og:image*. La imagen que representa al objeto dentro del Open Graph. (Obligatoria)
- *og:description*. Descripción del objeto.

Existen otras propiedades opcionales que, en nuestro caso, no se ajustan a nuestras necesidades. La información completa se encuentra en la propia documentación de Open Graph [98]. Como se puede ver en la Figura 72 estas propiedades se colocan en etiquetas <meta> en la cabecera (<head>) de la página web.

```
<meta property="og:type" content="article" />
<meta property="og:title" content="Multiscale Materials Modeling Group | UVA" />
<meta property="og:description" content="Website of the Multiscale Materials Modeling research
group of the University of Valladolid." />
<meta property="og:url" content="https://www.ele.uva.es/~mmm/" />
<meta property="og:image" content="https://www.ele.uva.es/~mmm/img/logo_mmm_rrss.png" />
```

Figura 72. Sintaxis de las etiquetas de Open Graph.

- Twitter Cards

Al añadir estas etiquetas veremos como aquellos Tweets en los que aparece un enlace a nuestro contenido reciben información complementaria a través de una “Card”, en la que aparece una imagen, un video, ... tal y como podemos ver en la Figura 73.



Figura 73. Ejemplo Twitter Card.

Existen varios tipos de “Cards”, en función del contenido que tratan: Summary, Summary con imagen, App y Player (Video, audio). En nuestro caso la que mejor se ajusta a nuestras necesidades es la “Summary”, que tiene las siguientes propiedades con su respectiva sintaxis:

- *twitter:card*. En este caso su valor debe ser “summary”. (Obligatoria)
- *twitter:title*. Título conciso relacionado con el contenido. (Obligatoria)
- *twitter:image*. URL de la imagen que representa el contenido.
- *twitter:description*. Breve descripción del contenido. (Obligatoria)

Existen propiedades que, en nuestro caso, no se ajustan a nuestras necesidades. La información completa se encuentra en la propia documentación de las Twitter Cards [99]. Como podemos ver en la Figura 74, y al igual que en el caso de las etiquetas del Open Graph, estas se sitúan en la cabecera (<head>) de la página, en etiquetas <meta>.

```
<meta name="twitter:card" content="summary"/>
<meta name="twitter:title" content="Multiscale Materials Modeling Group | UVa"/>
<meta name="twitter:description" content="Website of the Multiscale Materials Modeling research
group of the University of Valladolid."/>
<meta name="twitter:image" content="https://www.ele.uva.es/~mmm/img/logo_mmm_rrss.png"/>
```

Figura 74. Propiedades y sintaxis de las etiquetas de Twitter Cards.

Hemos empleado las etiquetas, tanto de Open Graph como de Twitter Cards, en todas las páginas del sitio web de forma que podamos tener cierto control en la forma en la que se muestran nuestras páginas cuando estas son compartidas a través de las redes sociales.

3.2.3.8. Otros

Se ha creado esta sección para tratar dos cuestiones que, si bien no tienen que ver con los factores relativos al código HTML, requieren el uso de etiquetas HTML y del tratamiento de metadatos. Además, dado que ambas cuestiones están relacionadas, pues tratan los datos bibliográficos de las publicaciones.

Se trata por un lado del marcado de datos que requiere Google Scholar, motor de búsqueda vertical que mencionamos previamente, para incluir las páginas web de las publicaciones. Por otro lado, y tal y como acordamos en la auditoría previa, procederemos a incluir y corregir, cuando sea pertinente, los datos bibliográficos ubicados en los metadatos de los archivos PDF de las publicaciones.

3.2.3.8.1. Google Scholar

Como hemos visto a lo largo de este documento, Google cuenta numerosos motores de búsqueda verticales, especializados en diferentes tipos de contenido como Google Imágenes, Google Books, Youtube...

Google Scholar ofrece al usuario una forma sencilla de buscar literatura académica de múltiples disciplinas y tipos, como pueden ser artículos, tesis, abstracts... obtenidos de múltiples fuentes, como editores académicos, sociedades académicas, repositorios online, universidades...

Desde la propia documentación de Google Scholar [\[100\]](#) se marcan una serie de directrices que los administradores web deben cumplir para un correcto rastreo e indexación de los documentos y su posterior visualización en los resultados de búsqueda. Las directrices están subdivididas en varias secciones que se corresponden con los estados del proceso de inclusión de un documento.

- Rastreo

Al igual que el buscador de Google “estándar”, Google Scholar emplea los robots o crawlers, que de forma automática rastrean la web y, en este caso, obtiene los documentos con el fin de incluirlos en el índice y mostrarlos en los resultados de búsqueda. Para ello los sitios web han de estar estructurados de forma que faciliten el trabajo de los robots. En concreto la estructura debe permitir que los crawlers sean capaces de descubrir y obtener todas las URLs correspondientes a los artículos, así como tener acceso para actualizar la información periódicamente.

Haremos un repaso pormenorizado de los elementos a tener en cuenta para un rastreo correcto:

- *Formato de los documentos*

Los archivos han de estar en formato HTML o en formato PDF. Los archivos PDF deben contar con texto “accesible”, es decir, deber ser posible poder buscar y encontrar palabras usando Adobe Acrobat Reader. Además, cada archivo no debe superar los 5 MB.

- *Accesibilidad*

Es necesario contar con una interfaz fácilmente navegable para que los robots de búsqueda encuentren las URLs de los artículos. Se recomienda que las URLs correspondientes a los artículos sean accesibles desde la página principal del sitio web a través de un máximo de 10 enlaces HTML.

Para el caso que nos ocupa en este proyecto, es decir, un sitio web con un número reducido de publicaciones, Google recomienda crear una lista con todas las publicaciones en un único archivo HTML, que incluya enlaces a los archivos PDF. Para sitios web con un mayor número de publicaciones (miles-millones) se recomienda proporcionar maneras de listar los archivos por fecha de publicación o de registro.

- *Disponibilidad del sitio web*

El sitio web debe permanecer accesible tanto para los usuarios como para los robots en todo momento. Los robots visitan las páginas periódicamente con el fin de actualizar contenido, así como para asegurarse de que las URLs están accesibles. Si los robots de búsqueda no pueden acceder a las URLs puede que los artículos desaparezcan de Google Scholar.

- **Indexación**

Google Scholar utiliza un software automatizado, denominado “parser”, capaz de identificar los datos bibliográficos de los documentos, así como referencias entre los ellos. La identificación incorrecta de estos datos provocará una mala indexación del documento. De hecho, debido a estos errores algunos documentos no pueden incluirse en el índice. Otros, en cambio, se incluirán con errores, pero estos provocarán una peor posición en las SERPs. Para evitar estos problemas se deben proporcionar los datos bibliográficos y las referencias de forma que el software automatizado pueda procesarlos.

- *Preparación previa de los artículos*

Cada artículo y cada abstract se colocará en archivos PDF y HTML separados. Por el momento no es posible indexar múltiples abstracts de una misma web o múltiples artículos de un mismo archivo PDF. Cada artículo deberá contar con su propia URL única si se desea que sea incluido en Google Scholar.

- *Meta-etiquetas*

Con el fin de permitir la correcta interpretación de los datos bibliográficos, Google Scholar permite el uso de meta etiquetas HTML. Para ello el sitio web deberá contar con una página web para cada una de las publicaciones, en esta página web se localizará el abstract correspondiente. De la misma manera esta página web enlazará con el archivo PDF de la publicación en cuestión.

Google Scholar soporta diferentes conjuntos de etiquetas, entre ellas, Highwire Press (ej. citation_title), Eprints (ej. eprints.title), BE Press (ej. bepress_citation_title) y PRISM (ej. prism.title). También soporta las etiquetas Dublin Core (ej. DC.title) pero Google no recomienda su utilización debido a que no cuenta con etiquetas que permitan identificar unívocamente algunos de los datos bibliográficos de publicaciones periódicas, como pueden ser el título de la publicación, el volumen o los números de página.

Además, existe la posibilidad de realizar esta tarea de una forma más o menos automática, para ello el sitio web deberá hacer uso de un repositorio o sistema de gestión de publicaciones, como Eprints, DSpace, Digital Commons o OJS. De ser así, bastará con configurarlo para que exporte los datos bibliográficos en <meta> etiquetas HTML.

A continuación, comentaremos las directrices para un correcto etiquetado y algunas de las etiquetas que son fundamentales a la hora de la indexación en Google Scholar. En nuestro proyecto emplearemos el conjunto de etiquetas de Highwire Press (ej. *citation_title*), pues es el que emplea Google Scholar en sus ejemplos y parece el más extendido entre los sitios web que cuentan con el meta etiquetado.

- Título (*citation_title*). Debe contener el título del artículo. No confundir con el título de la revista o libro en el que fue publicado. Esta etiqueta es necesaria para la inclusión del artículo en Google Scholar.
- Autor (*citation_author*). Debe contener el nombre de los autores reales del artículo. Esto no incluye a colaboradores, asesores... Cada nombre deberá figurar en una etiqueta diferente y siguiendo los siguientes formatos: "Santos, Iván" o "Iván Santos". Se necesita al menos una de estas etiquetas para la inclusión del artículo en Google Scholar.
- Fecha de publicación (*citation_publication_date*). Debe contener la fecha de publicación. Si es posible se proporcionarán fechas completas en el siguiente formato "aaaa/mm/dd", si no es posible bastará con el año de publicación. Esta etiqueta es necesaria para la inclusión del artículo en Google Scholar.
- Para los artículos de revistas y conferencias se emplearán las siguientes etiquetas para proporcionar los datos bibliográficos restantes: *citation_journal_title* o *citation_conference_title*, *citation_issn*, *citation_isbn*, *citation_volume*, *citation_issue*, *citation_firstpage* y *citation_lastpage*. Estas etiquetas deben contener información suficiente como para identificar una referencia a este artículo desde cualquier otro. Normalmente esta información es la siguiente: el nombre de la publicación o de la conferencia, el volumen, la cuestión (issue), y el número de la primera página del artículo en dicho volumen.
- Para tesis, disertaciones e informes técnicos se emplearán las siguientes etiquetas para proporcionar los datos bibliográficos restantes: *citation_dissertation_institution*, *citation_technical_report_institution* y *citation_technical_report_number*. Al igual que para los artículos de revistas y las conferencias, es necesario proporcionar datos suficientes para la identificación una referencia al artículo desde cualquier otro.
- Para todo tipo de documentos, la norma a seguir es presentar los artículos como normalmente se citaría en la sección "Referencias" de otro artículo. La omisión o la presentación incorrecta de datos bibliográficos puede ocasionar una mala identificación de los artículos.
- Las etiquetas son atributos HTML y por tanto se daba tratar los caracteres especiales de forma adecuada.
- Las etiquetas "<meta>" normalmente se aplican únicamente a la página web en la aparecen. Si en la página web únicamente se muestra el abstract del artículo y se cuenta

con la versión completa del mismo en un archivo independiente (PDF), será necesario especificar su localización mediante la etiqueta *citation_pdf_url*. El contenido de esta etiqueta deberá corresponderse con la URL absoluta del archivo PDF. En principio el archivo PDF deberá estar localizado en el mismo subdirectorio que el archivo HTML del abstract que le referencia. Un fallo a la hora de enlazar ambas versiones del documento puede derivar en problemas a la hora de indexar los archivos PDF.

```
<meta name="citation_title" content="Atomistic modeling of dopant implantation, diffusion, and activation" />
<meta name="citation_publisher" content="American Vacuum Society" />
<meta name="citation_doi" content="10.1116/1.2348726" />
<meta name="citation_language" content="en" />
<meta name="citation_author" content="Lourdes Pelaz" />
<meta name="citation_author" content="María Aboy" />
<meta name="citation_author" content="Pedro López" />
<meta name="citation_author" content="Luis A. Marqués" />
<meta name="citation_publication_date" content="2006" />
<meta name="citation_journal_title" content="Journal of Vacuum Science and Technology B" />
<meta name="citation_journal_abbrev" content="J. Vac. Sci. Technol. B" />
<meta name="citation_volume" content="24" />
<meta name="citation_issue" content="5" />
<meta name="citation_firstpage" content="2432" />
<meta name="citation_lastpage" content="2436" />
<meta name="citation_pdf_url" content="https://www.ele.uva.es/~mmm/papers/2006_Pelaz_JVSTB_24.pdf" />
```

Figura 75. Sintaxis de las etiquetas Highwire Press para Google Scholar.

La Figura 75 muestra un ejemplo del código correspondiente al etiquetado de una de las publicaciones. En nuestro ejemplo hemos empleado el conjunto de etiquetas de Highwire Press (ej. *citation_title*), pero independientemente del conjunto de etiquetas elegido es necesario proporcionar al menos tres campos: el título del artículo, el nombre completo de al menos uno de los autores, y la fecha de publicación. Las páginas que no proporcionen alguna de estas etiquetas se procesarán como si las etiquetas no existieran. Así mismo, todos los archivos PDF se procesan como si no tuvieran meta etiquetas, salvo que estos estén vinculados con su correspondiente página HTML que contiene su abstract, mediante la etiqueta *citation_pdf_url* anteriormente comentada. De esta forma proporcionamos la información bibliográfica de ambas versiones.

- Contenido sin meta-etiquetas

En el caso de que no se pueda implementar un sistema de meta etiquetado HTML, y los artículos solo están disponibles en formato PDF será necesario que el formato de estos se ajuste a las siguientes directrices:

- El título del artículo deberá ser la mayor porción de texto al comienzo de la página. El tamaño de la fuente deberá ser de al menos 24 pt. Será necesario emplear la misma fuente en todo el título. Debemos asegurarnos de que el resto del texto de la página tiene un tamaño de fuente inferior al del título o, por el contrario, este texto podría ser interpretado como el título del artículo.
- Los autores del artículo deberán figurar justo antes o justo después del título en un tamaño de fuente sensiblemente menor que el del título, pero mayor que el del texto normal. El tamaño de la fuente deberá ser de 16-23 pt. Al igual que en el título será necesario emplear la misma fuente para todos los autores. Debemos asegurarnos de que los nombres del repositorio y de la revista, así como los textos correspondientes a

los encabezados de las secciones, tengan un tamaño de fuente menor que el de los autores, a fin de evitar interpretaciones incorrectas.

- Incluir una cita bibliográfica de una versión publicada del artículo. Esta deberá figurar en el encabezado o en el pie de página de la primera página del archivo PDF.
- Evitar el uso de fuentes Tipo 3, suelen provocar errores en los tamaños de fuente y la codificación de caracteres, lo que dificulta el trabajo de los *parsers* que extraen los datos bibliográficos.

Una vez que se han aplicado todas las directrices anteriormente expuestas solo queda esperar para su indexación. Cuando se trata de artículos nuevos esto puede llevar días. Sin embargo, la actualización de artículos previamente indexados normalmente tarda del orden de 6-9 meses. En el caso de artículos alojados en sitios web con un amplio catálogo la actualización puede llevar años.

Hasta el inicio de este proyecto los artículos alojados en el sitio web del Multiscale Materials Modeling Group únicamente estaban disponibles en formato PDF, y por tanto su indexación en el motor de búsqueda Google (Google Scholar incluido) dependía únicamente de que este fuera capaz de identificar correctamente los datos bibliográficos, tal y como se ha expuesto anteriormente.

Como se vio en la Sección 3.1, correspondiente con la Auditoria previa, el estado de indexación de las publicaciones y el aspecto que estas tenían en las SERPs presentaban problemas. Esto, junto con lo visto en esta sección, pone de manifiesto que la identificación de algunos de los datos bibliográficos no es la correcta, lo que deriva en problemas de indexación y en errores a la hora de visualizar los resultados.

Por ello la mejora del estado de indexación de las publicaciones era uno de los principales objetivos de este proyecto. Una de las medidas que se tomaron fue la creación de páginas HTML para cada una de las publicaciones, con el fin de, entre otras cosas, tener un mayor control sobre los datos que se proporcionaban desde nuestro sitio web y con ello mejorar la indexación y el aspecto de nuestros artículos en las SERPs.

Con lo visto en esta sección, incluida la posibilidad de asociar la versión HTML que contiene el abstract junto con los datos bibliográficos y la versión completa en formato PDF, queda claro que ese es el camino a seguir para conseguir una indexación más precisa para nuestros artículos. Además de esta forma tenemos la posibilidad de indexar dos páginas para el mismo contenido aumentando así la posibilidad de una visita. Quizá no para los artículos ya indexados pues para ellos el trabajo ya está hecho, pero si para los artículos nuevos y aquellos que todavía no están incluidos en el índice.

3.2.3.8.2. Corrección de los datos bibliográficos de los archivos PDF

En un intento por solucionar los errores a la hora de mostrar correctamente los títulos de los artículos en formato PDF en las SERPs, se ha llevado a cabo un proceso de corrección de los datos bibliográficos que aparecen en los metadatos de los archivos en formato PDF.

Desde la propia documentación de Adobe [101] se indica que muchos motores de búsqueda utilizan el título que aparece en los metadatos a la hora de mostrar el documento en las SERPs. Si el archivo PDF careciera de título se mostraría el nombre del archivo, y este no tiene porqué corresponderse necesariamente con el título del mismo.

Estos datos pueden ser modificados desde la herramienta Adobe Acrobat, en la sección propiedades. En la Figura 76 podemos observar su formato:

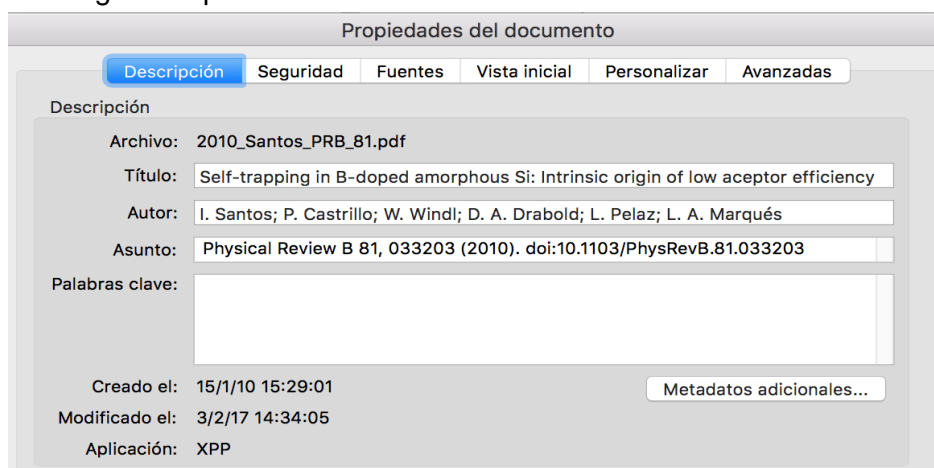


Figura 76. Metadatos/propiedades documento PDF.

En la revisión de los documentos nos encontramos con muchos casos en los que los datos son incorrectos, están incompletos o directamente no existen. Dado que no hay un formato estándar para completar estos datos, se ha elegido el formato de aquellos archivos PDF que presentaban todos sus datos de forma correcta. Como se puede ver en la imagen el archivo cuenta con un campo de título y de autor. En el campo asunto se hace referencia a la revista en la que el documento fue publicado, su volumen, su primera página y la fecha de publicación. Además, se incluye el DOI del documento.

En principio, y atendiendo a las recomendaciones de Google Scholar que vimos anteriormente, no parece que los metadatos del propio archivo PDF figuren entre las fuentes de las que el motor de búsqueda toma los datos bibliográficos para la identificación del archivo PDF. De hecho, la documentación de Google Scholar indica que los datos bibliográficos (título, autor, revista, volumen...) se toman del propio documento PDF en función de su formato.

Sin embargo, parece que para el motor de búsqueda de carácter general estos metadatos si que cobran importancia. Como se puede ver en el Figura 77, en el que se incluyó una errata en los metadatos del archivo PDF (“acceptor” en lugar de “acceptor”), parece que el motor de búsqueda toma el título del artículo de los metadatos del archivo PDF en lugar de obtenerlo del propio contenido del PDF.

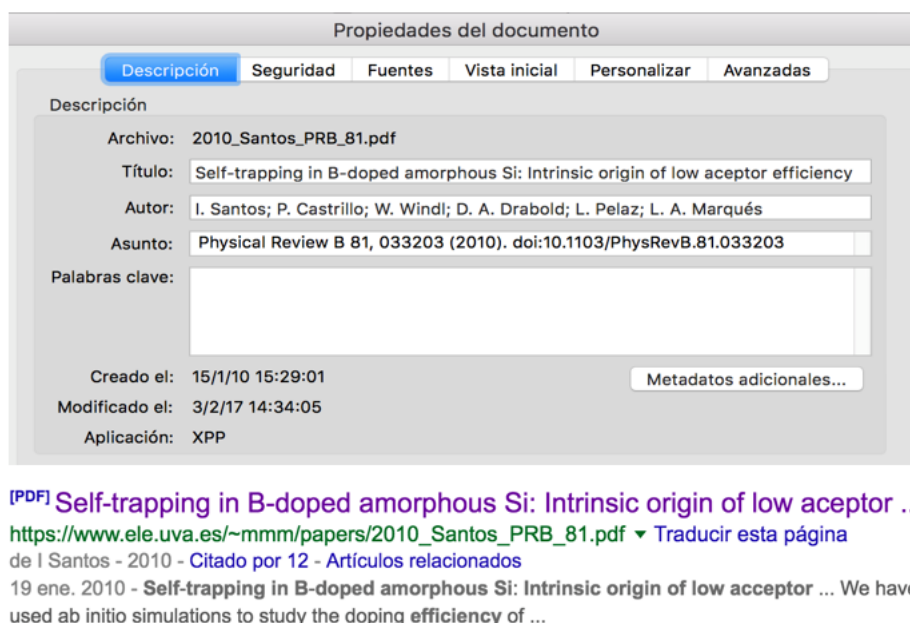


Figura 77. Ejemplo de la influencia de los metadatos PDF en la etiqueta de título en las SERPs.

Por tanto, queda demostrado que los metadatos de los archivos PDF tienen cierta influencia en el aspecto final que presentan en las SERPs, y por ello es recomendable prestar especial atención a que estos metadatos sean incluidos y que lo hagan de forma correcta, con el fin de evitar malas indexaciones o una mala imagen entre los usuarios al no representar fielmente el contenido del archivo.

3.3. SEO Off-page

Dado que no tenemos control sobre este tipo de factores nos limitaremos a hacer un breve inciso sobre algunos aspectos relativos a los enlaces y a las redes sociales.

3.3.1. Enlaces

Los enlaces son el principal factor externo que los motores de búsqueda tienen en cuenta a la hora de establecer la posición de una página web dentro del ranking. Como ya hemos visto, los motores de búsqueda emplean el análisis de los enlaces a lo largo de la web, comúnmente denominado Link Graph, para determinar la relevancia de los contenidos.

Como vimos en la Sección 2.1.2.2, existen una serie de factores que determinan la influencia que tendrán los enlaces recibidos sobre la posición de una página en los resultados de búsqueda. Generalmente, al tratarse de un factor externo, no tenemos ningún control sobre los enlaces que recibimos de otros sitios web. Aun así, existen una serie de técnicas y estrategias para la obtención de enlaces.

- Obtención de enlaces

Como hemos visto a lo largo de este proyecto los motores de búsqueda buscan que los enlaces se obtengan de forma natural, por ello algoritmos como Penguin persiguen el spam, la compra de enlaces, ... generalmente asociados a sitios web de baja calidad.

Aun así, existen diversas fuentes y formas de obtener enlaces. Todas ellas requieren un esfuerzo en la actividad de relaciones públicas, con el fin de promocionar el sitio web y su contenido. A continuación, analizaremos la situación algunas de ellas:

- *Periódicos*: Los periódicos son una de las fuentes de enlaces de mayor autoridad. La situación ideal pasa por recibir un enlace desde una noticia, pero salvo para grandes compañías resulta difícil. Antigüamente se hacía uso de las notas de prensa con el fin de obtener enlaces, pero dado que generalmente llevan asociado el pago de una cantidad económica no están bien vistas por los motores de búsqueda y pueden ser penalizadas.
- *Blogs*: Al igual que los periódicos, algunos blogs cuentan con una gran autoridad. Buscaremos los blogs más relevantes para el mercado en el que está ubicado el sitio web y trataremos de llegar a acuerdos de colaboración con sus autores. La colaboración comenzará por unos comentarios constructivos en las distintas entradas del blog, el reporte de errores y enlaces caídos, ... y terminará en una relación más estrecha con artículos invitados, consultoría técnica... a cambio de enlaces.
- *Artículos invitados*: Consiste en escribir un artículo de calidad en un blog, a poder ser con cierta reputación, a cambio de 1 o 2 enlaces a nuestro sitio web. De esta forma ganan las dos partes, el autor del blog obtiene contenido de calidad sin necesidad de generarlo y el autor del artículo obtiene enlaces de calidad.
- *Redes sociales*: A lo largo de los años ha cambiado la forma en la que los usuarios comparten el contenido. Actualmente las redes sociales están a la cabeza en este aspecto y, por tanto, son una gran fuente de enlaces. Hablaremos más de este tema en la Sección 3.3.2.
- *Blog propio*: Consiste en crear nuestro propio blog de referencia en el sector. En dicho blog crearemos contenido de calidad empleando las palabras clave para las que deseamos mejorar la posición y lo enlazaremos con la página web que deseamos mejorar. El problema es que requiere un gran esfuerzo y dedicación en la creación de contenido. Además, resultaría conveniente diversificar las fuentes del contenido enlazado, haciendo referencia a otros sitios web.

Como se puede ver la mayoría de técnicas de obtención de enlaces pasan por la promoción del sitio web, contactando con los autores de los distintos sitios web e intentando establecer acuerdos de colaboración, ya sea mediante artículos invitados, consultoría técnica, ... y por la generación de contenido de calidad adicional. En el caso que nos ocupa, y dado el propósito y la temática del sitio web objeto de este proyecto, no parece que ninguna de ellas se ajuste a nuestras posibilidades.

El primer lugar debemos destacar el propósito del sitio web. Este pertenece a un grupo de investigación, el Multiscale Materials Modeling Group, perteneciente a la Universidad de Valladolid, cuyos recursos están destinados a los diferentes proyectos que desarrollan y, por tanto, la generación de contenido se restringe a la publicación de los mismos. Los recursos materiales y personales dedicados a la generación de este contenido no son comparables a los empleados por prácticamente

cualquier blog, periódico... Por tanto, la generación de contenido para la obtención de enlaces no está entre sus objetivos.

Además, tenemos que una vez más la temática especializada reduce significativamente las alternativas para obtener enlaces de calidad. Y más si tenemos en cuenta que los motores de búsqueda consideran sospechosos aquellos enlaces que provienen de sitios web que no están relacionados con la temática del contenido enlazado. Las alternativas, por tanto, son bastante reducidas:

- *Redes de investigación:* Actualmente el sitio web recibe enlaces de redes de investigación españolas con la que colabora, como NanoSpain, Modelling for Nanotechnology y Variable Network.
- *Plataformas de literatura académica:* Nos referimos a plataformas como ResearchGate, Academia.edu, ScienceDirect... que nos permiten compartir los artículos académicos y cuyas funcionalidades sociales facilitan su divulgación.
- *Colaboradores:* Podemos obtener enlaces de sitios web pertenecientes a colaboradores, tanto personas como otros grupos de investigación. La incidencia de estos enlaces es relativa, pues generalmente se trata de enlaces de cortesía, es decir, ambos sitios web se enlazan recíprocamente. Los motores de búsqueda son capaces de detectar estos enlaces y reducen su importancia.

Por último, centraremos nuestro foco en el propio formato de nuestro contenido. Como hemos visto a lo largo de este proyecto, la práctica totalidad del contenido del sitio web, cerca del 90%, son los artículos académicos publicados por el grupo de investigación. Y esto supone un problema.

El sistema de citaciones de los artículos académicos hace prácticamente imposible la obtención de enlaces de forma natural, es decir, a través de citaciones en otros artículos. Cuando un artículo académico hace referencia a otro artículo lo hace, en la práctica totalidad de los casos, a través de su DOI (Document Object Identifier), un número específico que nos permite localizar el artículo a través de la Red.

Tomemos como ejemplo el DOI de uno de los artículos del grupo de investigación, y supongamos que únicamente contamos con dicho número para identificar el artículo. Tenemos dos opciones:

- Acudir a la página doi.org: nos permite buscar el documento asociado a dicho DOI.
- A través del enlace <http://doi.org/10.1063/1.3679126>

En ambos casos el DOI nos redirige a la página web del artículo correspondiente a la revista científica donde fue publicado.

Por tanto, el propio sistema establecido para la referenciación de los artículos anula en la práctica cualquier posibilidad de obtener enlaces a nuestros artículos académicos de la forma natural, es decir, a través de la referencia desde otros documentos.

3.3.2. Redes sociales

La reestructuración de la sección ‘Publications’ llevada a cabo en la sección 3.2.2.1, supuso la creación de una página web para cada una de las publicaciones del sitio web. Esto nos permite acabar con las limitaciones que nos imponían las versiones PDF de dichas publicaciones y añadir nuevas funcionalidades.

En concreto se han añadido botones que permiten compartir contenido a través de las redes sociales. Tal y como se puede observar en la Figura 78, en nuestro caso hemos decidido añadir botones para Facebook, Twitter, Google+, LinkedIn y Reddit. También se ha añadido la opción de compartir el contenido a través del correo electrónico.

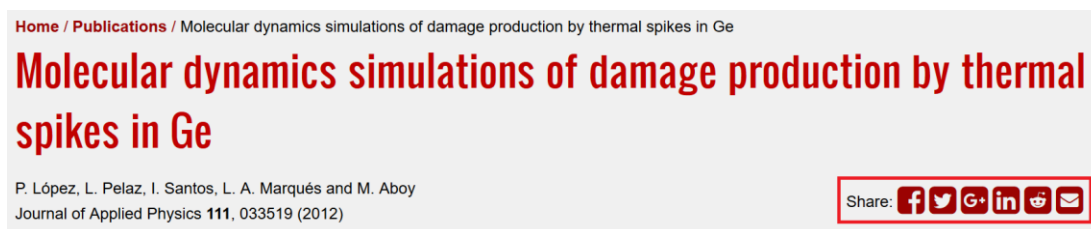


Figura 78. Ejemplo botones redes sociales.

Además, y tal y como vimos en la sección 3.2.3.7 se han añadido las etiquetas del Open Graph (Facebook) y las Twitter Cards, que nos permiten mejorar la forma en la que las redes sociales interpretan el contenido de las páginas enlazadas.

La evolución natural en este aspecto sería la apertura de cuentas en las diferentes redes sociales y mantener una actividad constante en las mismas, ya sea con la publicación de contenido o bien con interactuando con la comunidad asociada al sector, siguiendo a otros grupos de investigación, compartiendo información de conferencias, ...

Como podemos ver esta actividad requiere, salvando las distancias, la labor de un ‘Community manager’. Pero al igual que ocurría con las técnicas de obtención de enlaces, el esfuerzo, la dedicación y los recursos que requiere esta tarea, que en otro tipo de actividad sería indispensable, en nuestro caso no resulta abordable.

A diferencia de otros sitios web de carácter general, en los que esta labor puede ser llevada a cabo por personas no especializadas en la materia, en nuestro caso se requeriría de una persona cualificada y especializada en la materia. Pero esto no es posible, pues los recursos del grupo de investigación, tanto humanos como materiales, están destinados a los distintos proyectos que lleva a cabo.

Además, nos encontramos nuevamente con que la temática del sitio web reduce nuestras posibilidades. El público objetivo al que va dirigido el contenido es tan reducido y tan especializado que no compensa el esfuerzo de llevar a cabo esta tarea. En otra situación, con otro tipo de negocio, como por ejemplo un sitio web de noticias, una tienda online... donde el contenido es de carácter general y cualquiera puede ser un usuario potencial, la visibilidad en las redes sociales es indispensable, ya sea mediante campañas de publicidad en las mismas, contenidos virales...

Por último, debemos diferenciar entre los distintos caracteres de los sitios web. Sitios web como los descritos anteriormente (noticias, tiendas online...) tienen como objetivo la monetización del contenido,

ya sea a través de la publicidad o las ventas de artículos. Por tanto, los recursos empleados en visibilizar los contenidos a través de las redes sociales pueden ser mucho mayores. En nuestro caso el carácter del sitio web, meramente divulgativo y centrado en la investigación, hacen que los beneficios obtenidos no compensen la inversión en recursos necesaria para llevar a cabo esta actividad.

4. Seguimiento y análisis de los resultados.

Todos los cambios realizados sobre el sitio web del que es objeto este proyecto tienen como objetivo mejorar el posicionamiento de las páginas web del mismo en las SERPs de los motores de búsqueda.

Para comprobar el efecto que los distintos cambios provocan sobre el posicionamiento de las páginas web es necesario llevar a cabo un proceso de seguimiento de las palabras clave pertinentes. Para ello emplearemos la lista de palabras clave que obtuvimos en la Sección 3.2.1.2, correspondiente al proceso de Investigación de palabras clave (Sección 3.2.1.2) y que se encuentra en la Tabla A.2 del Anexo. También se llevará a cabo un seguimiento, algo menos extenso, de los títulos de las publicaciones con el fin de realizar búsquedas más específicas que nos permitan una mejor comprobación del estado de indexación de las publicaciones, tanto en su versión web como en PDF.

Además, se seguirá la evolución a lo largo del proyecto del estado de indexación global de las páginas del sitio web, para tratar de determinar qué modificaciones provocan mejoras en el rastreo e indexación del contenido.

Pero antes de empezar mencionaremos las herramientas empleadas para realizar este seguimiento y la metodología empleada en el mismo.

4.1. Herramientas

Existen multitud de herramientas que nos permiten llevar a cabo el seguimiento de un conjunto de palabras. La mayoría de estas herramientas son de pago, y cuentan con versiones de prueba gratuitas, limitadas en el tiempo y en sus funcionalidades. Dada la naturaleza de este proyecto, y a fin de mantener cierta coherencia y estabilidad en los resultados, se descartan este tipo de herramientas por su limitación temporal. Aun así, estas herramientas se han probado en paralelo con la herramienta finalmente elegida con el fin de comprobar la validez de los datos, y estos eran similares.

En nuestro caso se ha optado por emplear herramientas gratuitas o bien de pago, pero con versiones de prueba sin limitaciones temporales. Emplearemos las siguientes herramientas:

- *Google Search Console*: Las herramientas de administración de Google nos permiten controlar todos los aspectos de la administración de un sitio web. Una vez registrada la propiedad de un sitio web podemos ver el tráfico del sitio web, el estado de indexación de sus páginas, el marcado y detección de datos estructurados, eliminar URLs, subir sitemaps, consultar las estadísticas de rastreo, controlar los enlaces recibidos, dependiendo del tipo de dominio orientar a un país determinado... En nuestro caso, y dado que nuestro sitio web no es el dominio raíz, algunas de estas funcionalidades, como las estadísticas de rastreo, no son accesibles. De forma similar el hecho de que el dominio sea .es nos impide orientar el sitio web a un determinado país. Los datos proporcionados por la herramienta provienen exclusivamente de la navegación a través de Google.

El uso de la herramienta es gratuito y está supeditado a la posesión y registro en la misma de un sitio web.

- *Google Analytics*: Google Analytics nos ofrece información detallada del tráfico recibido por las páginas del sitio web,
- *Herramientas de administrador de Bing*: Al igual que las herramientas de administración de Google nos permiten controlar todos los aspectos de la administración del sitio web. Salvo aspectos puntuales, las funcionalidades ofrecidas son similares. En este caso los datos provienen de la navegación a través de Bing y Yahoo!. El uso de la herramienta es gratuito y está supeditado a la posesión y registro en la misma de un sitio web.
- *Rank tracker*: Para el seguimiento de las posiciones en las distintas palabras clave emplearemos la herramienta Rank tracker que forma parte del SEO PowerSuite, el paquete de herramientas comercializado por la empresa Link-Assistant.com. Estas herramientas cuentan con una versión gratuita sin límite de tiempo, aunque con limitación de algunas de sus funcionalidades, que se ajusta perfectamente a nuestras necesidades.

Entre las múltiples funcionalidades de Rank tracker se encuentra la posibilidad de realizar el seguimiento de las posiciones que ocupan en las SERPs las páginas de un sitio web ante una determinada consulta. La herramienta nos permite realizar el seguimiento de una lista de palabras clave a través de más de 400 motores de búsqueda, seleccionando el número de resultados a revisar, el idioma, el país...

Dada la ingente cantidad de datos recogidos durante el seguimiento se ha elaborado una pequeña herramienta que nos permita su visualización. La herramienta nos permite seleccionar una de las palabras clave de la lista y observar gráficamente su evolución en las posiciones a lo largo del proyecto tanto de forma global como mes a mes. La herramienta se encuentra en el disco (cuyo contenido figura en el Anexo) junto con los archivos de registro del seguimiento que contienen la información completa.

4.2. Metodología

Una vez que conocemos las herramientas que emplearemos a la hora de realizar el seguimiento describiremos la metodología empleada en el mismo.

Nuestro objetivo principal es realizar el seguimiento de las posiciones que ocupan en las SERPs las páginas de nuestro sitio web ante un conjunto de palabras clave, a fin de determinar cómo afectan las mejoras aplicadas a lo largo del proyecto a dichas posiciones. Realizaremos el seguimiento en varios motores de búsqueda de forma que podamos comparar la incidencia de las modificaciones en cada uno de ellos.

Los motores de búsqueda elegidos son Google, Google Mobile, Bing y Yandex. La elección de Google y Bing es obvia, pues se trata de los principales motores de búsqueda a nivel mundial. Elegimos Google Mobile con el objetivo de detectar la existencia de cambios en la posición de las páginas web frente a su versión de escritorio. En cuanto a Yandex, en un principio el objetivo era DuckDuckGo, pero dado que ninguna herramienta lo soportaba decidimos realizar el seguimiento de uno de sus suministradores de resultados, Yandex. Por el camino se quedan muchos motores de búsqueda, entre

de ellos algunos importantes, como Yahoo!. Esto se debe a Yahoo! comparte resultados con Bing y por tanto carece de sentido realizar el mismo trabajo dos veces.

Una vez elegidos los motores de búsqueda sobre los que se llevará a cabo el seguimiento es hora de establecer nuestro mercado objetivo. En este proyecto nuestro mercado objetivo es Estados Unidos, y será para las versiones de los motores de búsqueda de dicho país para las que se realizará el seguimiento. Periódicamente se seguirán las posiciones en la versión española de Google y Bing, a fin de determinar si el hecho de contar con un dominio .es hace que las distintas páginas web vean mejoradas sus posiciones frente a la versión de Estados Unidos.

Por último, cabe destacar la frecuencia del seguimiento. El hecho de llevar a cabo un seguimiento sobre una lista de palabras clave afectará negativamente a las métricas de uso de las páginas web que aparezcan en los resultados. Esto se debe a que el número de impresiones, es decir, el número de veces que una página aparece en los resultados de búsqueda se verá incrementado. Y dado que las impresiones producidas por el seguimiento no producen visitas, el CTR se verá reducido.

Por ello se deberá establecer una frecuencia de seguimiento responsable, que nos permita recoger un número de datos significativo y afectando lo menos posible a métricas de uso. En nuestro caso se han establecido varias frecuencias de seguimiento a lo largo del proyecto, reduciéndose paulatinamente a medida que el proyecto avanzaba y los resultados se iban asentando. También hay que tener en cuenta la carga de trabajo que supone el seguimiento y el tratamiento de los datos obtenidos del mismo.

Durante el primer mes se llevó a cabo un seguimiento diario, pero resultaba inviable dada la carga de trabajo que suponía. Durante los dos meses siguientes se pasó a realizar el seguimiento cada dos días. Posteriormente, y durante 11 meses, se realizó el seguimiento cada tres días. Por último, durante los dos últimos meses se pasó a realizar el seguimiento cada 10 días, con el fin de determinar si una vez que finalizará el seguimiento y las aguas volvieran a su cauce las posiciones se mantendrían.

En el caso del seguimiento en las versiones española de Google y Bing la frecuencia se estableció en 10 días durante todo el proceso de seguimiento.

A continuación, enumeramos los parámetros fundamentales del proceso de seguimiento:

- *País:* Estados Unidos (y periódicamente España).
- *Motores de búsqueda:* Google, Google Mobile, Bing y Yandex.
- *Número de resultados:* 120
- *Frecuencia:* Durante las primeras dos semanas se llevó a cabo un seguimiento diario. Posteriormente se pasó a realizar el seguimiento cada dos días durante algunos meses, para finalmente llevar a cabo el seguimiento cada tres días a fin de no saturar demasiado las palabras clave.
- *Duración:* 16 meses. (12 España)
- *Palabras clave:* Lista obtenida en el proceso de investigación de palabras clave. Ubicada en la Tabla A.2. del Anexo.

De forma similar se llevará a cabo un seguimiento de los títulos de las publicaciones. En este caso concentramos el análisis en el periodo de tiempo con mayor índice de indexación (Marzo 2017 - Junio 2017), con el fin de contar con los datos más completos. En este caso el seguimiento únicamente se llevará a cabo para Google.

4.3. Análisis de los resultados

A la hora de analizar los resultados existen varios enfoques. Desde una visión global en la que se analizan datos como la evolución en la aparición de resultados para el conjunto de palabras clave a seguir y la posición de estas a lo largo del tiempo, la evolución en la indexación de las páginas web del sitio web por parte de los motores de búsqueda... O bien desde una visión mucho más cercana, en la que se examinan la evolución de los resultados de palabras clave específicas.

4.3.1. Visión global

- Estados unidos

Los datos de la Tabla 7 reflejan la evolución en el número de palabras clave en las que aparece una de nuestras páginas web entre sus 120 primeros resultados. Como podemos ver, los datos están agrupados en función de la posición ocupada por la página. También se incluyen el número de papers (archivos PDF) y el número de publicaciones (archivo HTML), con el objetivo de observar el nivel de implantación de uno y otro a lo largo del tiempo. Los datos completos, incluyendo cada uno de los meses, se encuentran en la Tabla A.4 del Anexo.

Tabla 7. Evolución del posicionamiento.

	Marzo '16				Agosto '16				Enero '17				Junio '17			
	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM
1-3	2	1			5	4	5	5	5	4	2	5	5	5	2	5
4-10	2	3	-	-	6	6	4	6	6	5	5	8	9	5	4	9
11-20	1	1	-	-	9	3	4	8	12	3	1	12	7	3	1	7
21-50	7	3	-	-	10	13	3	12	13	13	6	15	8	7	4	9
51-120	11	4	-	-	17	10	3	14	17	6	3	13	16	8	4	15
Total	23	13	-	-	47	35	19	45	52	32	16	52	46	28	14	45
Papers	9	6	-	-	12	5	10	10	8	7	10	8	3	6	13	3
Publi.	0	0	-	-	1	11	3	1	10	5	1	10	16	6	0	15

Analizaremos estos datos en función del motor de búsqueda, haciendo hincapié en el caso de Google:

- Google

Google muestra una mejora constante en los resultados obtenidos a lo largo del proyecto, tanto en el número de palabras clave que cuentan con alguna de nuestras páginas entre sus resultados y la posición que ocupan dichas páginas, como en el número de publicaciones que aparecen en los resultados. Por tanto, parece que las modificaciones introducidas han funcionado en mayor o menor medida. A continuación, detallaremos las cifras de estas mejoras.

Al inicio del proyecto, solamente 23 de las 180 palabras clave de la lista contaban con una de nuestras páginas web entre sus resultados, de las que únicamente 4 estaban situadas en la primera SERP.

A lo largo del tiempo la situación ha mejorado, rondando de forma constante las 50 palabras clave en las que aparecía alguna de nuestras páginas e incluso superando dicha cifra en varias ocasiones. En Noviembre de 2016 y en Enero de 2017 llegó a su momento álgido con un total de 52 palabras clave en las que aparecía alguna de nuestras páginas, logrando un aumento del 126% respecto a la situación inicial.

Cabe señalar que, aunque el mayor número de palabras clave con resultados sea 52, dada la volatilidad de algunos de los resultados el número de palabras clave que han recibido resultados en algún momento del proyecto es mayor, cerca de las 70.

En cuanto al número de páginas que aparecen en la primera SERP nos encontramos que la cifra ha ido aumentando a lo largo del proyecto, superando de forma constante la decena. El mejor momento se dio en Octubre de 2016 y en Junio de 2017 con un total de 14 palabras clave con resultados en la primera página, logrando un aumento del 250% respecto a la situación inicial.

Por último, nos detendremos en el número de publicaciones que aparecen en los resultados, tanto en su versión PDF (papers) como en su versión HTML (publicaciones). En un principio únicamente 9 de las publicaciones aparecían en los resultados de alguna de las palabras clave de la lista, y como es lógico todas ellas en su versión PDF. Una vez introducidas las versiones HTML (Abril de 2016) de cada una de las publicaciones la situación mejora, obteniendo en Noviembre de 2016 su mejor dato, con 20 publicaciones en los resultados, de las que 13 son archivos PDF y 7 archivos HTML. Esto supone un aumento del 122%.

Ese mismo mes se introducen las secciones de 'Publicaciones relacionadas', con el objetivo de mejorar el estado de indexación de las publicaciones. A partir de ese momento la situación comienza a revertirse, y los archivos HTML de las publicaciones comienzan a superar a los archivos PDF. En Junio de 2017 contamos con 19 publicaciones en los resultados de las que 16 son archivos HTML y 3 son archivos PDF. En general, las versiones HTML de las publicaciones obtienen mejores posiciones. Esto nos confirma la conveniencia de contar con estas versiones para complementar las versiones PDF.

- *Bing*

En el caso de Bing la situación al inicio del proyecto es algo peor. Solamente 13 de las 180 palabras clave de la lista contaban con una de nuestras páginas web entre sus resultados, de las que únicamente 4 estaban situadas en la primera SERP.

La situación mejora hasta llegar a contar con 36 palabras clave en las que aparecen alguna de nuestras páginas web, lo que supone un aumento del 177%. Pero, aun así, lejos de las 52 que llegó a tener Google. Además, nos encontramos con que en el caso de Bing la situación es más volátil, y los resultados obtenidos no se mantienen tanto en el tiempo.

En cuanto a las páginas que aparecen en la primera SERP, la cifra aumenta hasta las 10 en varios puntos del proyecto. Esto supone un incremento del 150% frente a las 4 páginas iniciales.

Por último, nos detendremos en el número de publicaciones que aparecen en los resultados. En un principio contábamos únicamente con 6 publicaciones que aparecían en los resultados, todas ellas en su versión PDF. Tras la introducción de las versiones HTML (Abril 2016) de cada una de las publicaciones, en Agosto de 2016 obtenemos el pico máximo de 16 publicaciones entre los resultados, con 5 archivos PDF y 11 archivos HTML. Tras este pico máximo, la situación a lo largo del proyecto se estabiliza en las 12 publicaciones. Al contrario que ocurre con Google, la introducción de las secciones de 'Publicaciones relacionadas' no surten ningún efecto.

- *Yandex*

El caso de Yandex es paradójico. Alterna constantemente meses con buenos resultados con meses nefastos, por lo que no parece que los cambios hayan afectado a los niveles de indexación ni a las posiciones. Aun así, parece que es el motor de búsqueda que mejor interpreta los archivos PDF, pues la mayoría de los resultados obtenidos se corresponden con dichos archivos.

- *Google Mobile*

Las cifras de Google Mobile, tanto en el número de palabras clave que cuentan con alguna de nuestras páginas entre los resultados y la posición que ocupan dichas páginas, como en el número de publicaciones que aparecen en los resultados, son similares a las de su versión de escritorio, y por tanto no entraremos en detalle. Sí cabe destacar que en general y a lo largo de todo el proyecto los resultados de Google Mobile obtienen resultados sensiblemente mejores que en su versión de escritorio, alrededor de 1-3 posiciones.

- **España**

Como hemos comentado previamente, el objetivo de llevar a cabo un seguimiento en las versiones españolas de Google y Bing no es otro que tratar de determinar la influencia que tiene el hecho de contar con un dominio de tipo ccTLD. Este tipo de dominios indican que el contenido de dicho sitio web está dirigido a un único país. En nuestro caso contamos con un dominio .es por lo que nuestro contenido está dirigido a España.

La Tabla A.5, ubicada en el Anexo, nos muestra los datos de seguimiento de las versiones españolas de Google y Bing.

Como podemos ver Google mantiene unas cifras y una evolución similar a la de su versión americana a lo largo de todo el proyecto. Únicamente se observa cierta mejora en la posición de algunas de las páginas, como la página principal y las páginas de la sección 'Research', llegando a tener 20 palabras clave con nuestras páginas web situadas en la primera SERP. Por tanto, parece que, en el caso de Google, el dominio de tipo ccTLD únicamente aporta cierta mejora en la posición de determinadas páginas sin afectar al grado de indexación. Algo que sí ocurre con Bing con veremos a continuación.

En el caso de Bing la situación no puede ser más diferente que la de su versión americana. Se trata de hecho la versión que mejores resultados globales obtiene de todos los motores de búsqueda sobre los que se ha llevado a cabo el seguimiento.

De media supera ampliamente las 60 palabras clave con alguna de nuestras páginas entre los resultados, llegando a un máximo de 69 en Noviembre de 2016. Este dato por sí solo demuestra que, en el caso de Bing, el dominio de tipo ccTLD resulta determinante en la indexación de nuevas páginas.

También resulta determinante en las posiciones ocupadas por estos resultados, pues nos muestran una situación única. Una gran parte de los resultados obtenidos se encuentran en la primera SERP. De los 69 resultados obtenidos en Noviembre de 2016, 52 estaban situados en la primera página.

Algo similar ocurre con las publicaciones, especialmente en sus versiones HTML. En su punto álgido, Octubre de 2016, aparecían en los resultados un total de 36 publicaciones, más de un tercio del total (91). De estas 29 corresponden a las versiones HTML de las mismas. Lo que termina de confirmar la conveniencia de contar con estas versiones para complementar las versiones PDF.

4.3.2. Estado de indexación

- Estado de indexación general

En la Sección 3.1, correspondiente con la Auditoría previa, analizamos el estado de indexación del sitio web al comienzo del proyecto. Dicho análisis nos mostró que existían ciertos problemas para la indexación de los archivos PDF correspondientes a las publicaciones. A lo largo del proyecto se han llevado a cabo una serie de modificaciones que tiene como objetivo, entre otras cosas, solucionar dichos problemas. Se trata de modificaciones como la inclusión de página HTML para cada una de las publicaciones, las secciones de 'Publicaciones relacionadas', el HTTPS...

En la Figura 79 tenemos una gráfica que nos permite visualizar cómo afectan las modificaciones al estado de la indexación. El eje vertical se representa el número de páginas indexadas, mientras que el eje horizontal representa el tiempo. La grafica representa cuatro series de datos:

- *Número de páginas*: Representa el número total de páginas web incluidas en el proyecto.
- *Sitemap*: Representa las páginas web indexadas que están incluidas en el archivo Sitemap.xml. Su número máximo se corresponde con *Número de páginas*.
- *HTTP*: Representa las paginas indexadas con el protocolo HTTP. Su número máximo es superior al marcado por *Número de páginas*, pues se incluyen páginas que si bien están alojadas en el sitio web no formaban parte del proyecto.
- *HTTPS*: Representa las paginas indexadas con el protocolo HTTPS. Al igual que en el caso anterior su número máximo es superior al marcado por *Número de páginas*.

Como podemos ver en la Figura 79 existe un punto, tras la introducción del protocolo HTTPS, en el que se produce la migración de HTTP a HTTPS. El estado de indexación de HTTP comienza a caer mientras que el HTTPS aumenta.

Por último, las líneas verticales marcan la fecha en la que fueron incluidas aquellas modificaciones susceptibles de provocar un cambio en el estado de indexación.

Los hitos principales en la mejora del estado de indexación corresponden a la introducción de las nuevas páginas de las secciones 'Research', 'Publications' y 'Projects'. Como es lógico todas ellas llevan asociado un aumento en el número máximo de páginas del sitio web y un incremento en el número de páginas indexadas. En el caso de las páginas de las publicaciones podemos observar que, tras un buen inicio, cerca de dos meses los niveles de indexación caen.

Otra de las modificaciones que produce cambios profundos en el estado de indexación es la migración a HTTPS. Como podemos observar conlleva la re-indexación de todas las páginas del sitio web. Si bien el número total de páginas indexadas se mantiene durante el proceso, conviviendo las versiones HTTP y HTTPS, se necesita de al menos 3 meses para que el estado de indexación de las páginas HTTPS alcancen niveles semejantes a los que tenían las páginas HTTP.

Por último, tenemos la introducción de las secciones de 'Publicaciones relacionadas'. Esta modificación provoca un gran aumento en los niveles de indexación, logrando las mejores cifras de indexación de todo el proyecto, con un 95% de las páginas indexadas. Además, esta mejora en los niveles de indexación se mantiene durante meses.

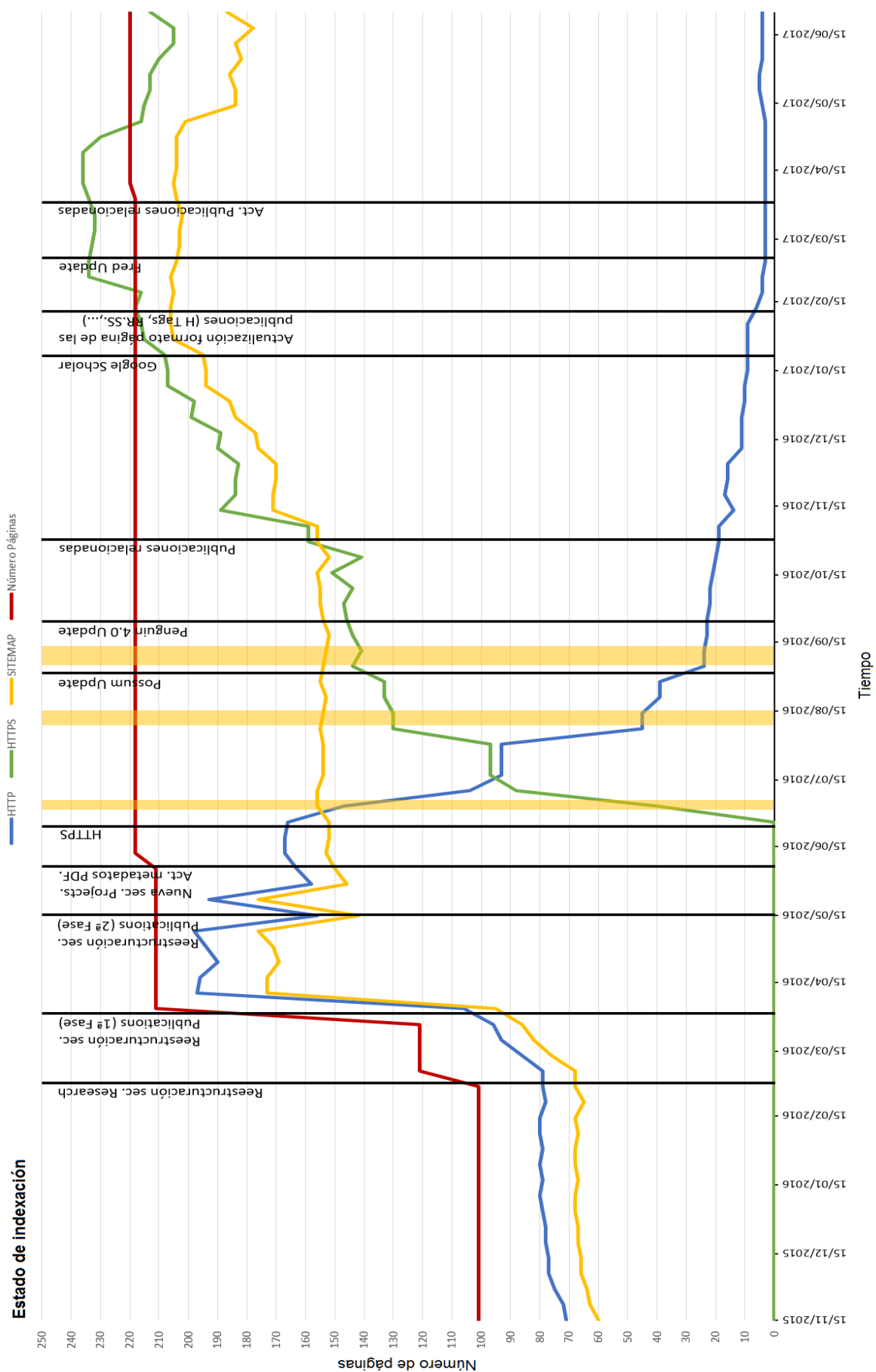


Figura 79. Estado de indexación.

- Análisis de los títulos de las publicaciones

En la sección 3.1.2, correspondiente a la Auditoria previa, se comprobó que el estado de indexación de las publicaciones, en ese momento asociado únicamente a los archivos PDF, dejaba bastante que desear. Entre un 20-40% de las publicaciones no aparecían indexadas, y algunas de las que si lo estaban presentaban serios problemas en el formato en que aparecían en las SERPs debido a la mala interpretación de sus datos bibliográficos.

Para tratar de solucionar estos problemas se tomaron dos medidas. Por un lado, la revisión y corrección de la información bibliográfica ubicada en los metadatos de los archivos PDF. Por otro lado, creando páginas web para cada una de las publicaciones que nos permitieran tener un mayor control sobre el aspecto que las publicaciones presentan en las SERPs.

Como hemos comentado previamente en la sección 4.2, se ha llevado a cabo un seguimiento de los títulos de las publicaciones durante el periodo con mayor grado de indexación (Marzo '17 – Junio '17) con cerca del 95% de las paginas indexadas (206/220 páginas). El fin es comprobar que los motores de búsqueda identifican correctamente las publicaciones y las asocian unívocamente con sus respectivos títulos. Se han comprobado tanto los resultados que aparecen visible como los resultados ocultos, ya que la alta cantidad de versiones de un mismo documento hace que los motores de búsqueda solo muestren aquellos que consideren más relevantes ocultando el resto.

Tabla 8. Análisis de los títulos de las publicaciones.

	Marzo '17	Abril '17	Mayo '17	Junio '17
PDF	52	63	59	51
HTML	76	90	84	76
Solo PDF	11	0	3	8
Solo HTML	35	26	28	33
PDF+HTML	41	63	57	43

La Tabla 8 muestra los datos relativos a las versiones PDF y HTML. Comenzaremos con los archivos PDF. Con cerca del 95% de las páginas del sitio web indexadas presenta en su mejor momento un total de 63 publicaciones identificadas e indexadas correctamente por su título. Este supone cerca de un 69% del total de publicaciones, por lo que parece que los problemas a la hora de reconocer los datos bibliográficos persisten aun logrando buenas cifras de indexación.

En cuanto a los archivos HTML se ha llegado a contar con un 100% de publicaciones correctamente identificadas e indexadas y las cifras generalmente no han bajado del 80%. Estos datos demuestran la eficacia que ha supuesto la introducción de las versiones HTML para cada una de las publicaciones, permitiéndonos ganar visibilidad en las SERPs.

La Tabla 8 también nos muestra el número de publicaciones que cuentan únicamente con la versión PDF, con la versión HTML y con las dos versiones. Esto nos permite ver ahondar en la importancia de contar con las versiones HTML pues complementan perfectamente a los PDF y suplen sus carencias.

4.3.3. Casos concretos

Una vez analizada la situación global, en este apartado trataremos casos concretos, relativos a la evolución en las posiciones que siguen las páginas web ante una determinada palabra clave. En general trataremos casos que nos permitan ver la influencia concreta de algunas de las modificaciones aplicadas al sitio web. Además, trataremos casos especiales como el de la página principal, las nuevas páginas de la sección 'Research', las nuevas páginas de las publicaciones...

Dada la ingente cantidad de datos y el tamaño limitado de este documento nos limitaremos a mostrar una serie limitada de ejemplos. Para analizar todos los resultados recomendamos hacer uso de la herramienta creada específicamente para la visualización de los datos de seguimiento.

- Página principal

Nuestro principal objetivo en este proyecto es mejorar la posición de la página principal del sitio web, especialmente en Google. Se considera que la palabra clave representativa del sitio web y para la cual deseamos un mejor posicionamiento es '*multiscale materials modeling*'. La Figura 80 nos muestra la evolución en la posición de la página principal del sitio web a lo largo del proyecto para dicha palabra clave. La gráfica incluye, mediante líneas verticales, las principales modificaciones realizadas sobre el sitio web, lo cual nos permite relacionar la inclusión de estas modificaciones con las mejoras en las posiciones.

Como podemos ver en la Figura 80, la posición que ocupa la página principal en los cuatro motores de búsqueda para los que realizamos el seguimiento sufre una continua fluctuación. Especialmente en los casos de Bing y Yandex, que en ningún momento parecen mostrar cierta estabilidad, empeorando los resultados e incluso llegando a desaparecer, lo que indica que las modificaciones introducidas no surten ningún efecto.

Por el contrario, y pese a las fluctuaciones, las dos versiones de Google muestran una clara mejoría a lo largo del proyecto y unos resultados mucho más estables. La mejora en la posición implica una mejor reacción ante las modificaciones introducidas.

Originalmente la página principal ocupaba posiciones comprendidas entre la tercera y la cuarta SERP. Tras un mal inicio, en el que se pierden muchas posiciones, la situación mejora ostensiblemente. A partir de Julio de 2016 y tras la introducción del protocolo HTTPS, con su correspondiente re-indexación, la página principal empieza a ocupar posiciones comprendidas entre la primera y la segunda SERP. Aunque todas las modificaciones tienen su importancia en la posición de una página web, en nuestro caso, parece que la introducción del protocolo HTTPS marca la diferencia.

Como podemos ver el objetivo de mejorar la posición de la página principal se ha cumplido con creces. Especialmente en esos periodos en los que la página ha ocupado puestos en la primera SERP, llegando a estar situada por momento en la 5ª posición, codeándose con rivales difíciles de superar como Wikipedia, la página oficial de las conferencias MMM, ... y situándose como el primer grupo de investigación que aparece en los resultados. Hubiera sido deseable una mayor estabilidad en las apariciones en la primera SERP, pero aun así parece que la página ha encontrado la estabilidad en torno a las posiciones 10-14.

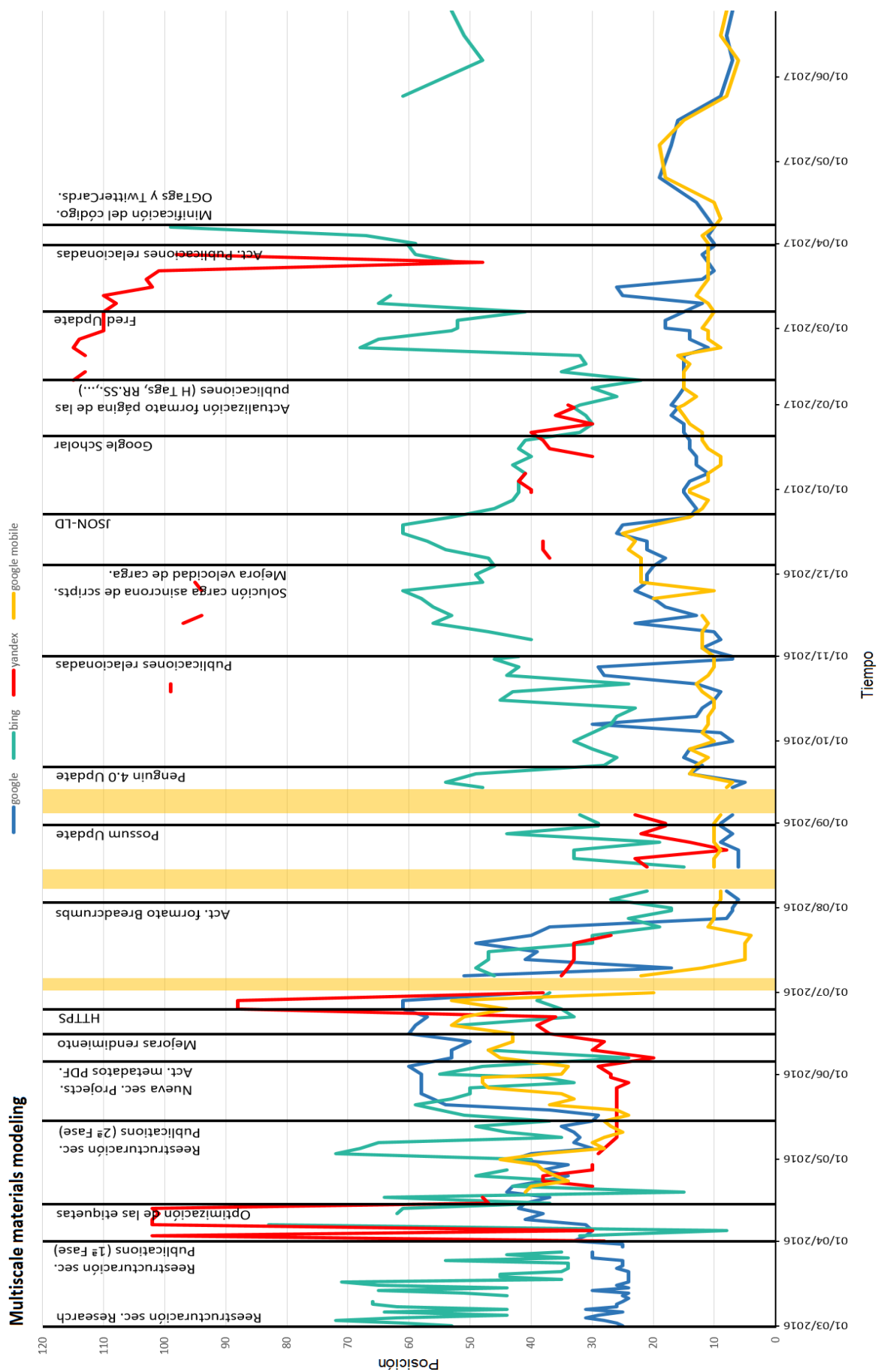


Figura 80. Evolución 'multiscale materials modeling'.

- Páginas de la sección 'Research'

Una de las modificaciones llevadas a cabo fue la reestructuración de la sección 'Research'. El objetivo era hacer un mejor uso de las palabras clave con las que contábamos, y por ello se crearon páginas web para cada uno de los temas tratados en las subsecciones de la sección 'Research'.

Los resultados obtenidos en cada una de estas páginas no pueden ser más dispares. Nos encontramos que algunas páginas obtienen resultados espectaculares (Figura 81) mientras que otras obtienen resultados mediocres o ni siquiera aparecen entre los resultados (Figura 82). Dado que todas las páginas han sufrido las mismas modificaciones y, por tanto, se encuentran en las mismas condiciones, todo parece indicar que la diferencia la marcan la competencia de las palabras clave para las que clasifican, y sobre todo el contenido de la página web en cuestión.

En la Figura 81 podemos ver la evolución en la posición de la página web correspondiente al apartado '*Molecular Implants*' [\[102\]](#), para la palabra clave '*molecular implants*'. Originalmente, cuando este contenido no contaba con su propia página web, ocupaba posiciones entre la quinta y la sexta SERP de Google. Tras la creación de la página web específica, comienza a aparecer de forma constante en la primera SERP de Google, y durante gran parte del proyecto ocupando el primer lugar.

Los cambios parecen tener una mayor influencia en los motores de búsqueda de Google que en Bing y Yandex. En este caso conseguimos posicionar la página en la primera SERP tanto para Bing como para Yandex, pero nuevamente nos encontramos con una falta de consistencia importante, con desapariciones que duran meses.

La Figura 81 únicamente muestra uno de los ejemplos. De forma similar, las páginas web de los apartados '*Damage generation mechanisms*', '*Description irradiation damage*' y '*Influence of irradiation damage*' logran copar de forma constante la primera posición de las SERPs. Existen otras páginas que, si bien no logran alcanzar la primera posición de las SERPs como en los casos anteriores, sí que alcanzan posiciones notables, a caballo entre la primera y la segunda SERP. Se trata de las páginas web de los apartados '*Binary collision approximation*', '*Influence irradiation parameters*' y '*Energetic characterization*'. Para visualizar la evolución en la posición de dichas páginas recomendamos hacer uso de la aplicación creada para tal cometido.

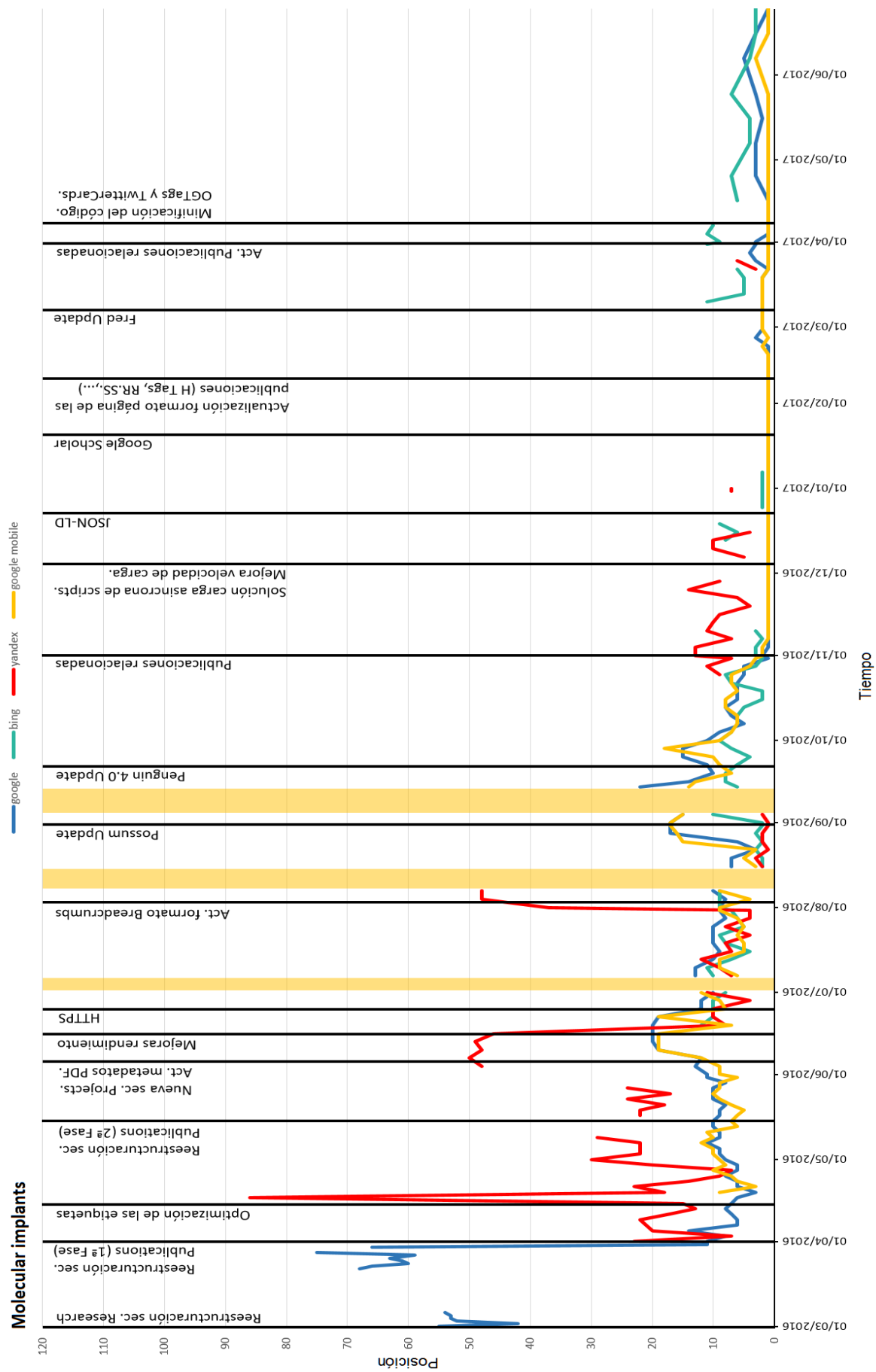


Figura 81. Evolución 'molecular implants'.

Por otro lado, tenemos casos en los que, si bien hemos conseguido que los motores de búsqueda lleguen a indexar las nuevas páginas para las palabras clave correctas, las posiciones obtenidas no son tan positivas ni tan constantes a lo largo del tiempo como en los casos anteriores. Generalmente ocupan posiciones irrelevantes, que van desde la cuarta hasta la décima SERP. La importancia radica en el hecho de que hagan acto de presencia entre los resultados, más allá de su posición, pues originalmente dichos contenidos no estaban indexados para sus correspondientes palabras clave.

La Figura 82 muestra la evolución en la posición de una de esas páginas, la correspondiente al apartado '*Classical Molecular Dynamics*' [\[103\]](#) y para la palabra clave del mismo nombre. Originalmente, cuando este contenido no contaba con su propia página web, no aparecía entre los resultados. Tras la creación de la página web específica su aparición entre los resultados no es inmediata. Tarda casi cuatro meses en hacer acto de presencia, y casi siete en hacerlo de forma continua. Como podemos ver, en el caso de Google, ocupa posiciones situadas entre la sexta y la octava SERP

En cuanto a la reacción ante las modificaciones incluidas, esta vez a Google se le une Bing. Como podemos ver en la Figura 82, los resultados de Bing muestran cierta constancia en sus apariciones en los resultados, pero no así en las posiciones que ocupan, que presentan fluctuaciones mucho más pronunciadas que en el caso de Google. Yandex no hace acto de presencia en este caso.

Nuevamente, la Figura 82 solo muestra un ejemplo. Nos encontramos con una situación similar en las páginas web de los apartados '*Amorphous films*', '*Tight-binding*', '*Dynamical studies*', '*Defect engineering*', '*Semiconductor doping*', '*Nanodevices*' y '*Kinetic monte carlo*'. Para visualizar la evolución en la posición de dichas páginas recomendamos hacer uso de la aplicación creada para tal cometido.

Por último, tenemos casos para los que las páginas web de nueva creación no han logrado su principal cometido, aparecer en las SERPs. Se trata de las páginas web de los apartados '*Ab initio*', '*Thermal treatments*', '*Photoluminescence*', '*Electronic properties*', '*Structural characterization*' y '*Sputtering*'.

Existen varios motivos por los que dichas páginas web no aparecen entre los resultados. El primero de ellos, la relevancia del contenido. Los motores de búsqueda pueden considerar que dicho contenido no resulta relevante para las palabras clave a las que está dirigido. También debemos tener en cuenta una peculiaridad que comparten algunas estas páginas web. Se trata de páginas que hacen uso de palabras clave con un mayor grado de competencia, pues tienen múltiples acepciones y se emplean en diferentes contextos y, por tanto, lograr posicionar una página para dichos términos resulta mucho más difícil.

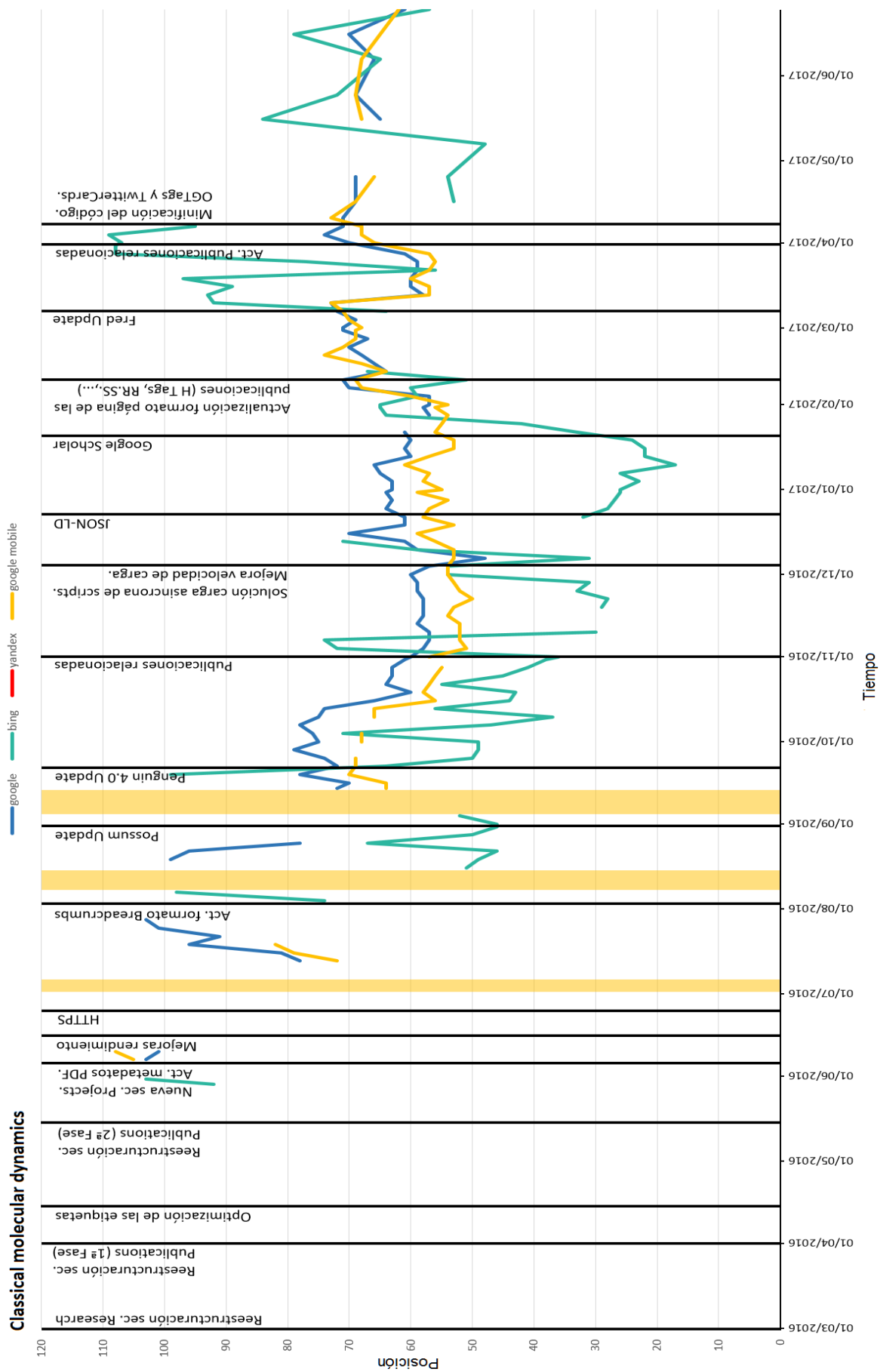


Figura 82. Evolución 'classical molecular dynamics'.

Como ya hemos comentado previamente las condiciones de partida de todas las páginas web de la sección 'Research' son las mismas. Por tanto, todo parece indicar que los factores determinantes a la hora de establecer la posición de dichas páginas serán el contenido de cada una de ellas y la competencia existente para las palabras clave en cuestión.

En el caso de Google, y a través de los diferentes casos que hemos visto hemos podido constatar que el contenido marca la diferencia. Mientras que en algunos casos logramos alcanzar de forma constante la primera SERPs e incluso copar la primera posición, en otros no somos capaces de obtener cierta constancia en los resultados y las posiciones obtenidas son francamente mejorables.

En cuanto a los datos de Bing y Yandex, especialmente este último, la falta de reacción ante los cambios introducidos en algunos casos y la inestabilidad de los resultados obtenidos en otros, no nos permiten sacar muchas conclusiones más allá de la evidente diferencia de resultados entre estos motores de búsqueda y Google.

- Páginas de la sección 'Publications'

En el análisis de los datos obtenidos del seguimiento de los títulos de las publicaciones hemos podido ver que las versiones HTML tienen un mejor desempeño que las versiones PDF. En ese caso se trataba de búsqueda específicas, empleando los títulos completos de las publicaciones. En este apartado veremos algún ejemplo de publicaciones que han aparecido en los resultados de búsqueda de algunas de las palabras clave de la lista.

Antes de pasar a los ejemplos hay que destacar que, en general, la introducción de las versiones HTML conlleva que las publicaciones aparezcan entre los resultados de una forma más constante. Especialmente en los dos motores de búsqueda de Google, que nuevamente muestran signos de una mejor recepción de las modificaciones relativas a las publicaciones. En casos puntuales Bing y Yandex dan muestra de cierta reacción ante los cambios, pero en general sus resultados y la constancia de los mismos no son comparables a los obtenidos por Google. Además, estos ejemplos nos permitirán ver algunos aspectos relativos al contenido duplicado y a la canibalización de palabras clave sobre los que se advirtió en la Sección 3.2.2.6 y la Sección 3.2.1.2.1 respectivamente.

La Figura 83 muestra la evolución en la posición de los resultados obtenidos para la palabra clave '*atomistic origin*'. Tras la introducción de la versión HTML, tenemos un periodo de tiempo en el que se obtienen ciertos resultados, pero nuevamente nos encontramos con una situación en la que tardamos meses en conseguir que nuestra página aparezca de forma constante en las SERPs. En concreto en la segunda SERP.

En este caso los resultados se corresponden únicamente con la versión HTML, lo que nos muestra la importancia de contar con dichas versiones, pues de otra forma esa publicación no aparecería en las SERPs para dicha palabra clave.

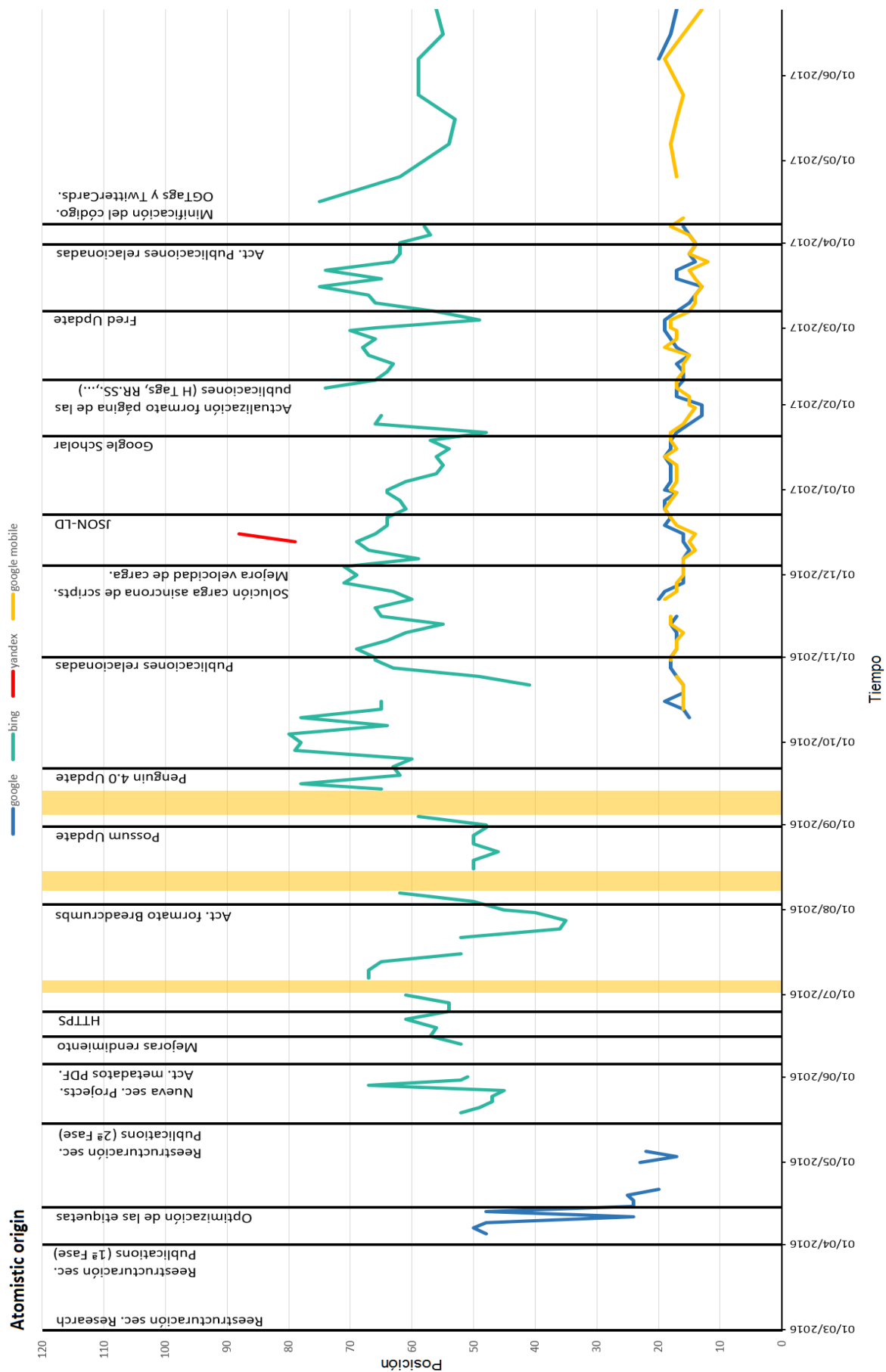


Figura 83. Evolución 'atomistic origin'.

La Figura 84 muestra la evolución en la posición de los resultados obtenidos para la consulta '*amorphous pockets*'. Centrándonos en Google, podemos ver tarda casi seis meses, desde introducción de las versiones HTML de las publicaciones, en aparecer de forma constante entre los resultados. Tras esto los resultados alternan periodos de constancia, en los que se logran posiciones en la primera SERP, con periodos de desaparición.

Durante el primer periodo de constancia en los resultados, que transcurre en los meses de Octubre, Noviembre y Diciembre, estos se corresponden con la versión PDF. Durante el segundo periodo, que comienza en Febrero de 2017 y continúa hasta el final del proyecto, los resultados se corresponden con la versión HTML de la publicación. No parece que la versión HTML nos dote con una mejor posición, pero sí que nos permite obtener una mayor constancia a lo largo del tiempo.

Entre esos dos periodos de constancia en los resultados nos encontramos un periodo en el que los resultados, tanto la versión PDF como la versión HTML, desaparecen. Durante ese periodo las SERPs muestran entre sus resultados versiones de la misma publicación, ubicadas en algunos de los sitios web que aloja el dominio www.ele.uva.es, perteneciente al Departamento de Electricidad y Electrónica de la Universidad de Valladolid, y que también aloja el sitio web del que es objeto este proyecto.

Se trata de un claro ejemplo de las consecuencias que acarrea contar con contenido duplicado dentro de un mismo dominio, tal y como advertimos cuando tratamos los factores SEO relativos a la Arquitectura de un sitio web, en la Sección 3.2.2.6. En este caso con mayor gravedad pues al tratarse de sitios web distintos dentro de un mismo dominio nuestras páginas desaparecen. Si el contenido duplicado estuviera restringido a nuestro sitio web, solo se mostraría una de las alternativas nuestras páginas no desaparecerían de las SERP por completo.

Por último, la Figura 85 muestra la evolución en la posición de los resultados obtenidos para la palabra clave '*preamorphized silicon*'. La peculiaridad de los resultados de esta palabra clave no está asociada a las posiciones obtenidas ni a la constancia de las mismas. Tampoco al hecho de que todos los motores de búsqueda para los que se realiza el seguimiento muestran buenos resultados.

La peculiaridad la marca el hecho de que cuatro de las publicaciones hacen referencia a la palabra clave '*preamorphized silicon*', y durante todo el proyecto han estado intercambiando el privilegio de aparecer en las SERP. Se trata de un claro caso de canibalización de palabras clave, en la que varias páginas de un sitio web clasifican para una misma palabra clave y los motores de búsqueda únicamente muestran la página que consideran más relevante.

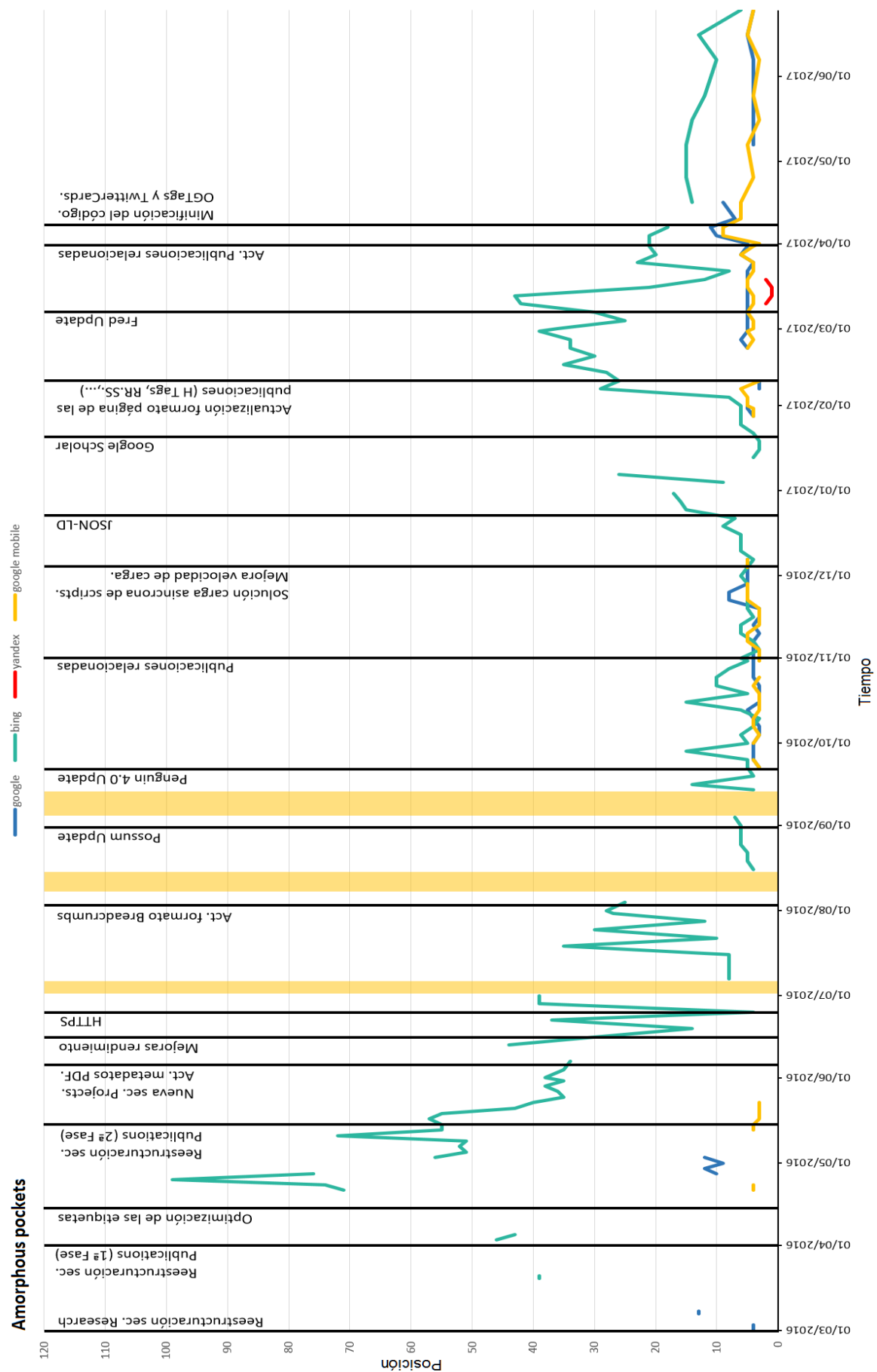


Figura 84. Evolución 'amorphous pockets'.

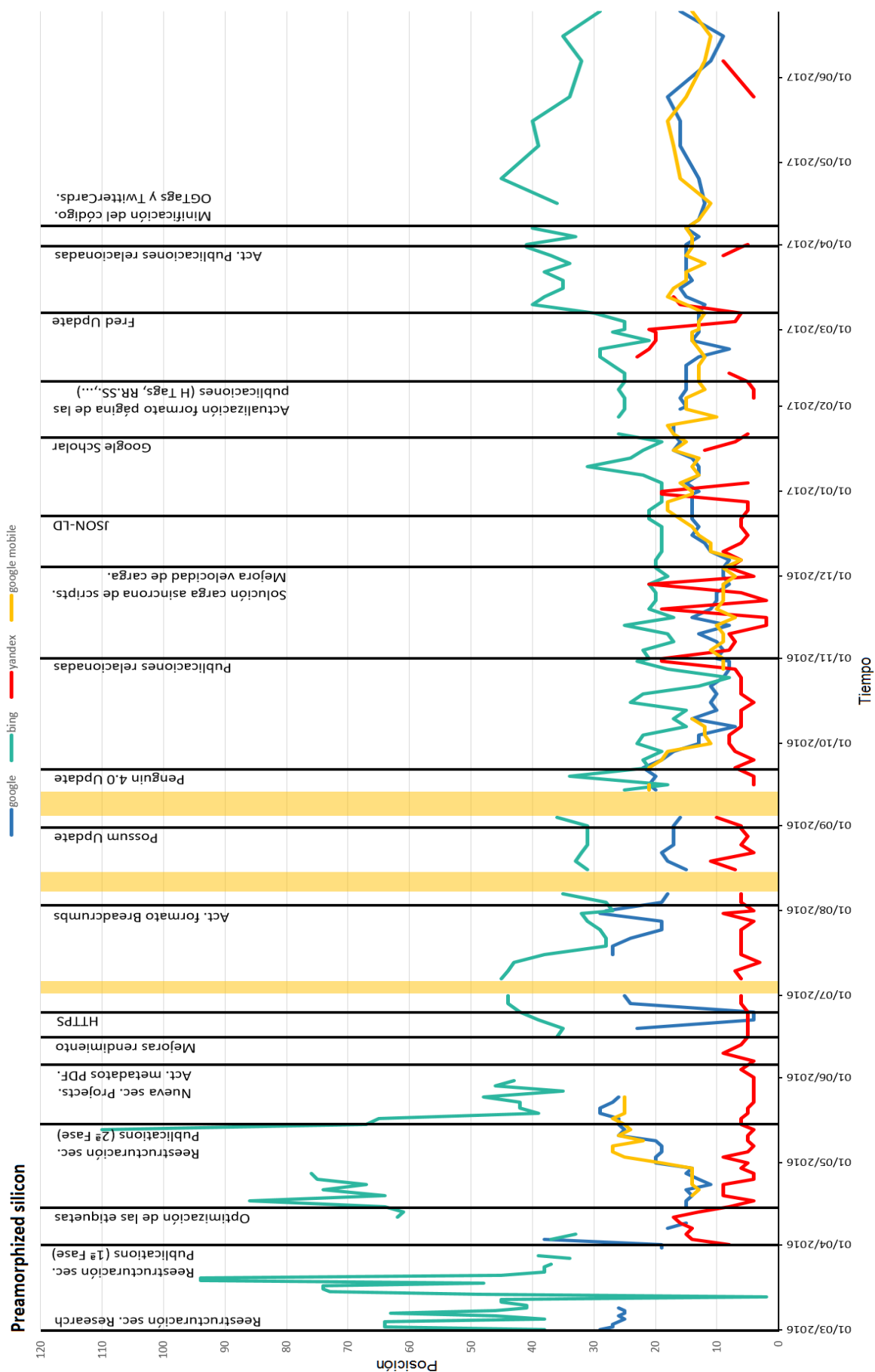


Figura 85. Evolución 'premorphized silicon'.

5. Conclusiones

A lo largo de este documento hemos descrito los principales factores que emplean los motores de búsqueda a la hora de posicionar una página web en los resultados de búsqueda. Uno de nuestros principales objetivos al inicio de este proyecto era intentar establecer la influencia de cada uno de estos factores en la posición de las páginas web, y para ello hemos llevado a cabo un seguimiento de las modificaciones realizadas sobre el sitio web objeto de este proyecto.

Como hemos visto resulta muy complicado determinar la influencia exacta de cada uno de los factores, por no decir que es imposible. Aun así, el análisis de los resultados obtenidos nos deja datos muy interesantes y clarificadores de la influencia de algunos de los factores. A fin de mantener la coherencia con el resto del documento dividiremos las conclusiones en función de la tipología de los factores en cuestión.

- **Contenido**

Hemos confirmado algo que, por otra parte, ya conocíamos: el contenido es el rey. Tanto los factores intrínsecos al contenido (calidad, frescura y unicidad), como sus factores asociados (investigación palabras clave, uso de las palabras clave, contenido duplicado) han resultado ser de vital importancia. Como hemos visto en la Sección 4.3, en igualdad de condiciones entre las distintas páginas web la calidad del contenido marca la diferencia. También es necesario destacar la importancia de la Investigación de palabras clave, pues nos ha permitido conocer las palabras clave empleadas en el sitio web y, a través de dicho conocimiento, hacer un uso más eficiente de las palabras clave con las que contamos.

- **Arquitectura y aspectos técnicos**

Los factores relativos a la arquitectura del sitio web han demostrado tener un gran peso sobre el posicionamiento de las páginas de nuestro sitio web. La introducción de una arquitectura optima con la reestructuración de las secciones 'Research' y 'Publications', estrechamente relacionado con la investigación y el uso de las palabras clave, nos ha permitido hacer un uso más eficiente de las palabras clave con las que contamos. También nos ha permitido establecer una jerarquía mucho más clara entre los distintos contenidos con los que contamos, algo en lo que ha ayudado la introducción de los Breadcrumbs. Por otra parte, la introducción de los apartados de 'Publicaciones relacionadas' nos ha permitido relacionar aquellos contenidos que compartían una temática, lo que nos ha servido para cohesionar el contenido del sitio web y para lograr alcanzar las mayores cotas de indexación gracias a una mejor distribución del PageRank.

Destacar también, la importancia de una correcta gestión del contenido duplicado interno. En nuestro caso la existencia de contenidos duplicados dentro del dominio, no de nuestra página web, ha provocado que algunos de nuestros resultados no hayan tenido las posiciones ni la continuidad deseada

En cuanto a los aspectos técnicos asociados a la arquitectura, la influencia de la optimización para dispositivos móviles ha quedado demostrada, pues generalmente los resultados obtenidos en Google Mobile son sensiblemente mejores que los de la versión de escritorio. En cuanto al

protocolo HTTPS, el salto de calidad en las posiciones obtenidas por la página principal tras su introducción demuestra la gran influencia que posee actualmente este factor.

- Código HTML

La influencia ejercida por factores relativos al código HTML resulta difícil de medir. Los distintos elementos asociados a los factores del Código HTML se encuentran incluidos en todas las páginas de nueva creación y, por tanto, contribuyen en la gran mejora detectada tras la creación de las nuevas páginas de las distintas secciones, pero esta importancia resulta difícilmente cuantificable. Además, se trata de elementos que tienen una estrecha relación con la usabilidad y la experiencia de usuario, y dado el reducido tráfico que recibe el sitio web resulta muy difícil medir su influencia mediante las métricas de uso.

En el caso de los datos estructurados, se ha comprobado positivamente la funcionalidad ofrecida por el marcado de los Breadcrumbs. En cuanto al resto de elementos marcados, no parece que hayan ejercido ninguna influencia sobre la posición de las páginas en las que están ubicados. Algo similar ha ocurrido con el marcado de la información bibliográfica para Google Scholar, aunque en este caso, con los tiempos que maneja dicho buscador vertical, la situación puede mejorar en el futuro.

La Tabla 9 muestra los distintos factores considerados en este proyecto, a los que se les ha asignado una puntuación en función de la importancia que consideramos que han tenido en nuestro proyecto. Los signos (+) indican el grado mayor o menor de relevancia y los signos (?) nos indican que se trata de factores que no se han considerado en este proyecto. Nos centramos únicamente en los factores On-Page pues son sobre los que tenemos control.

Tabla 9. Factores SEO y su relevancia.

On-Page						Factores Negativos	
Contenido		Arquitectura y aspectos técnicos		Código HTML			
Calidad (+++)	Uso palabras clave (++)	Arquitectura óptima (+++)	Tiempo de carga (+)	Etiqueta de título (+++)	Anchor text (++)		
Investigación keywords (+++)	Contenido vertical (+)	Formato URLs (+)	HTTPS (+++)	Etiqueta description (++)	Datos estructurados (++)		
Frescura (++)	Contenido duplicado ext. (+++)	Dispositivos móviles (+++)	Contenido duplicado int. (+++)	Etiquetas encabezado (+)	Open graph / Twitter Cards (+)		
				Etiqueta alt (+)			
Off-Page							
Confianza		Enlaces		Personalización		Redes sociales	
Autoridad (?)		Calidad (?)	Posición / Contexto (?)	Geo targeting (?)		Reputación (?)	
Historial (?)		Relevancia temática (?)	Velocidad de adquisición (?)	Resultados locales (?)		Número (?)	
Datos de uso (?)		Anchor text (?)	Nollow (?)	Historial (?)			
		Número (?)					

- Thin content
- Cloaking
- Keyword Stuffing
- Contenido oculto
- Piratería
- Anuncios abusivos
- Enlaces de pago
- Spam

En cuanto a la influencia que las modificaciones han tenido en los distintos motores de búsqueda, tenemos que destacar los resultados de Google frente a sus rivales, Bing y Yandex. Google ha

demostrado ser mucho más sensible a los cambios, tanto en la detección y la aceptación de los mismos, como en la mejora de los resultados tras la introducción de dichos cambios.

Para concluir haremos una reflexión sobre lo que ha supuesto este proyecto. Al comienzo del proyecto destacamos la naturaleza cambiante del SEO, en constante evolución. En mi experiencia yo destacaría otros elementos como la opacidad que en muchos momentos envuelve al proceso, lo que provoca momentos de confusión.

Esta sensación de opacidad se ve incrementada por el escaso control que tenemos sobre los resultados. La existencia de múltiples factores, algunos de ellos desconocidos para nosotros, hace que cada caso sea especial, pues las técnicas aplicadas sobre una página web no tienen porque funcionar en otra página. Por tanto, no se trata de un sistema determinista en el que tras la aplicación de una serie de técnicas obtengamos siempre los mismos resultados.

También hay que destacar que son los motores de búsqueda los que marcan en todo momento los tiempos del proyecto. Generalmente las modificaciones son detectadas en un tiempo relativamente corto, pero en otros casos se han necesitado periodos de tiempo mucho mayores para su detección. Situaciones como esta condicionan la correcta evolución del proyecto y nos lleva a introducir retrasos en la aplicación de otras modificaciones.

Por último, destacar la constante sensación de inestabilidad más allá de la constante fluctuación natural de los resultados de búsqueda. Puede darse el caso de que se hayan aplicado las modificaciones que permitan optimizar la posición de una página web, que dichas modificaciones hayan funcionado correctamente logrando mejorar sustancialmente la posición de la página web y que, tras un periodo de tiempo y sin llevar a cabo ninguna modificación, repentinamente dicha página desaparezca de los resultados de búsqueda y no vuelva a aparecer.

La lectura es clara, el hecho de haber obtenido una mejora sustancial en los resultados tras el proceso de optimización no nos garantiza que dichos resultados vayan mantenerse a lo largo del tiempo. Por tanto, tras la finalización de este proyecto recomiendo llevar a cabo un proceso de seguimiento tanto de la situación del sitio web como de las principales novedades que el mundo del SEO nos depare en el futuro, y en función de estas, llevar a cabo las modificaciones que sean pertinentes para mantener el sitio web en un estado óptimo.

6. Bibliografía

- [1] Enge E., Spencer S., y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition.
- [2] Barry Schwartz, "Google Made 890 Improvements To Search Over The Past Year", Search Engine Land, 19 Agosto 2014, <http://searchengineland.com/google-made-890-improvements-search-past-year-201065> (Última visita: 09/07/2017)
- [3] Google, "How search works. Useful responses", <https://www.google.com/search/howsearchworks/responses/> (Última visita: 09/07/2017)
- [4] Google Webmaster Central Blog, "What Crawl Budget Means for Googlebot", 16 Enero 2017, <https://webmasters.googleblog.com/2017/01/what-crawl-budget-means-for-googlebot.html> (Última visita: 09/07/2017)
- [5] Enge E., Spencer S., y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition, pp. 78-80.
- [6] Google, "How search works. Crawling and indexing", 24 Junio 2017, <https://www.google.com/search/howsearchworks/crawling-indexing/> (Última visita: 09/07/2017)
- [7] Google Official Blog, "Our new search index: Caffeine", 8 Junio 2010, <https://googleblog.blogspot.com.es/2010/06/our-new-search-index-caffeine.html> (Última visita: 09/07/2017)
- [8] Rand Fishkin, "Does Query Deserves Diversity Algorithm Exist at Google?", 11 Mayo 2008, <https://moz.com/blog/does-query-deserves-diversity-algorithm-exist-at-google> (Última visita: 09/07/2017)
- [9] Google, "How search works. Search Algorithms", 24 Junio 2017, <https://www.google.com/search/howsearchworks/algorithms/> (Última visita: 09/07/2017)
- [10] Enge E., Spencer S., y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition, pp. 56-70.
- [11] Danny Sullivan, "Google Launches "Universal Search" & Blended Results", Search Engine Land, 16 Mayo 2007, <http://searchengineland.com/google-20-google-universal-search-11232> (Última visita: 09/07/2017)
- [12] Peter J. Meyers, "Mega-SERP: A visual guide to Google", Moz, 10 Octubre 2013, <https://moz.com/blog/mega-serp-a-visual-guide-to-google> (Última visita: 09/07/2017)
- [13] Enquiro, "Google Golden Triangle Research: Eye Tracking Study", 2005, <http://go.pardot.com/l/22222/2013-04-24/qch> (Última visita: 09/07/2017)
- [14] Mediative, "The Evolution of Google's Search Engine Results Pages and Their Effects on User Behaviour", <http://www.mediative.com/whitepaper-the-evolution-of-googles-search-results-pages-effects-on-user-behaviour/> (Última visita: 09/07/2017)
- [15] Red Cardinal, "SERP Click Through Rate of Google Search Results - AOL data", 2006, <http://www.redcardinal.ie/google/12-08-2006/clickthrough-analysis-of-aol-datatz/> (Última visita: 09/07/2017)
- [16] Enquiro, "Business to Business 2007", 2007, <http://web.archive.org/web/20120131103323/http://www.enquiroresearch.com/campaigns/Business%20to%20Business%20Survey%202007.pdf> (Última visita: 09/07/2017)
- [17] Slingshot, "Google & Bing CTR study", 2011, <http://connect.relevance.com/a-tale-of-two-studies-establishing-google-bing-click-through-rates> (Última visita: 09/07/2017)
- [18] Chitika, "The Value of Google Result Positioning", 2013, <https://chitika.com/google-positioning-value> (Última visita: 09/07/2017)
- [19] Catalyst Digital, "How user intent impacts Google Click-Through Rates", 2013, <https://www.catalystdigital.com/wp-content/uploads/GoogleCTRStudy-Catalyst.pdf> (Última visita: 09/07/2017)

- [20] Caphyon, "Google Organic CTR Study 2014", <https://www.advancedwebranking.com/google-ctr-study-2014.html> (Última visita: 09/07/2017)
- [21] Philip Petrescu, "Google Organic Click-Through Rates in 2014", Moz, 1 Octubre 2014, <https://moz.com/blog/google-organic-click-through-rates-in-2014> (Última visita: 09/07/2017)
- [22] StatCounter, <http://gs.statcounter.com/> (Última visita: 09/07/2017)
- [23] Danny Sullivan, "Yahoo & Google Together Again In New Search Deal", Search Engine Land, 20 Octubre 2015, <http://searchengineland.com/yahoo-google-search-deal-233963> (Última visita: 09/07/2017)
- [24] DuckDuckGo, "Sources", <https://duck.co/help/results/sources> (Última visita: 09/07/2017)
- [25] Moz, "Google Algorithm Change History", Última visita: 25 Junio 2017, <https://moz.com/google-algorithm-change> (Última visita: 09/07/2017)
- [26] Masha Maksimava, "Google's 9 Major Algorithm Updates", 11 Abril 2017, <https://www.link-assistant.com/news/google-algorithm-updates.html> (Última visita: 09/07/2017)
- [27] Sergey Brin, Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", <http://infolab.stanford.edu/~backrub/google.html> (Última visita: 09/07/2017)
- [28] Danny Sullivan, "What Social Signals Do Google & Bing Really Count?", Search Engine Land, 1 Diciembre 2010, <http://searchengineland.com/what-social-signals-do-google-bing-really-count-55389> (Última visita: 09/07/2017)
- [29] Google Official Blog, "Introducing schema.org: Search engines come together for a richer web", 2 Junio 2011, <https://googleblog.blogspot.com.es/2011/06/introducing-schemaorg-search-engines.html> (Última visita: 09/07/2017)
- [30] Paul Breummer, "Future SEO; Understanding Entity Search", Search Engine Land, 7 Octubre 2013, <http://searchengineland.com/future-seo-understanding-entity-search-172997> (Última visita: 09/07/2017)
- [31] Alexis Sanders, "What Is Semantic Search and What Should You Do About It?", Moz, 14 Noviembre 2016, <https://moz.com/blog/what-is-semantic-search> (Última visita: 09/07/2017)
- [32] Google Official Blog, "Making search more secure", 18 Octubre 2011, <https://googleblog.blogspot.com.es/2011/10/making-search-more-secure.html> (Última visita: 09/07/2017)
- [33] Thom Craver, "Goodbye, Keyword Data: Google Moves Entirely to Secure Search", Search Engine Watch, 23 Septiembre 2013, <https://searchenginewatch.com/sew/news/2296351/goodbye-keyword-data-google-moves-entirely-to-secure-search> (Última visita: 09/07/2017)
- [34] Google Official Blog, "Introducing the Knowledge Graph: things, not strings", 16 Mayo 2012, <https://googleblog.blogspot.com.es/2012/05/introducing-knowledge-graph-things-not.html> (Última visita: 09/07/2017)
- [35] Google Webmaster Central Blog, "HTTPS as a ranking signal", 6 Agosto 2014, <https://webmasters.googleblog.com/2014/08/https-as-ranking-signal.html> (Última visita: 09/07/2017)
- [36] Google Official Blog, "Giving you fresher, more recent search results", 3 Noviembre 2011, <https://googleblog.blogspot.com.es/2011/11/giving-you-fresher-more-recent-search.html> (Última visita: 09/07/2017)
- [37] Google Official Blog, "Finding more high-quality sites in search", 24 Febrero 2011, <https://googleblog.blogspot.com.es/2011/02/finding-more-high-quality-sites-in.html> (Última visita: 09/07/2017)
- [38] Barry Schwartz, "Google Panda Is Now Part Of Google's Core Ranking Signals", Search Engine Land, 12 Enero 2016, <http://searchengineland.com/google-panda-is-now-part-of-googles-core-ranking-signals-240069> (Última visita: 09/07/2017)

- [39] Enge E., Spencer S., y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition, pp. 598-609.
- [40] Google Webmaster Central Blog, "Page layout algorithm improvement", 19 Enero 2012, <https://webmasters.googleblog.com/2012/01/page-layout-algorithm-improvement.html> (Última visita: 09/07/2017)
- [41] Google Official Blog, "Another step to reward high-quality sites", 24 Abril 2012, <https://webmasters.googleblog.com/2012/04/another-step-to-reward-high-quality.html> (Última visita: 09/07/2017)
- [42] Enge E., Spencer S., y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition, pp. 610-612.
- [43] Google Official Blog, "Penguin is now part of our core algorithm", 23 Septiembre 2016, <https://webmasters.googleblog.com/2016/09/penguin-is-now-part-of-our-core.html> (Última visita: 09/07/2017)
- [44] Google Disavow, <https://www.google.com/webmasters/tools/disavow-links-main> (Última visita: 09/07/2017)
- [45] Removing Content From Google, <https://support.google.com/legal/troubleshooter/1114905> (Última visita: 09/07/2017)
- [46] Google Inside Search Blog, "An update to our search algorithms", 10 Agosto 2012, <https://search.googleblog.com/2012/08/an-update-to-our-search-algorithms.html> (Última visita: 09/07/2017)
- [47] Barry Schwartz, "Google Payday Loan Algorithm: Google Search Algorithm Update To Target Spammy Queries", Search Engine Land, 11 Junio 2013, <http://searchengineland.com/google-pay-day-loan-algorithm-google-search-algorithm-update-to-target-spammy-queries-162941> (Última visita: 09/07/2017)
- [48] Danny Sullivan, "FAQ: All About The New Google Hummingbird Algorithm", Search Engine Land, 26 Septiembre 2013, <http://searchengineland.com/google-hummingbird-172816> (Última visita: 09/07/2017)
- [49] Greg Sterling, "It's Official: Google Says More Searches Now On Mobile Than On Desktop", Search Engine Land, 5 Mayo 2015, <http://searchengineland.com/its-official-google-says-more-searches-now-on-mobile-than-on-desktop-220369> (Última visita: 09/07/2017)
- [50] Google Webmaster Central Blog, "Finding more mobile-friendly search results", 22 Abril 2015, <https://webmasters.googleblog.com/2015/02/finding-more-mobile-friendly-search.html> (Última visita: 09/07/2017)
- [51] Barry Schwartz, "Now we know: Here are Google's top 3 search ranking factors", Search Engine Land, 24 Marzo 2016, <http://searchengineland.com/now-know-googles-top-three-search-ranking-factors-245882> (Última visita: 09/07/2017)
- [52] Danny Sullivan, "Schmidt: Listing Google's 200 Ranking Factors Would Reveal Business Secrets", Search Engine Land, 27 Septiembre 2010, <http://searchengineland.com/schmidt-listing-googles-200-ranking-factors-would-reveal-business-secrets-51065> (Última visita: 09/07/2017)
- [53] Danny Sullivan, "Dear Bing, We Have 10,000 Ranking Signals To Your 1,000. Love, Google", Search Engine Land, 11 Noviembre 2010, <http://searchengineland.com/bing-10000-ranking-signals-google-55473> (Última visita: 09/07/2017)
- [54] Rand Fishkin, "The beginner's SEO guide", Moz, 18 Diciembre 2015, <https://moz.com/beginners-guide-to-seo> (Última visita: 09/07/2017)
- [55] Google, "Google's SEO Starter Guide", <https://static.googleusercontent.com/media/www.google.com/es/webmasters/docs/search-engine-optimization-starter-guide.pdf> (Última visita: 09/07/2017)

- [56] Search Engine Land, "The Periodic Table Of SEO Success Factors", Junio 2017, <http://searchengineland.com/seotable/> (Última visita: 09/07/2017)
- [57] Search Engine Land, "Search Engine Land's Guide to SEO", Junio 2017, <http://searchengineland.com/guide/seo> (Última visita: 09/07/2017)
- [58] Moz, "Search Engine Ranking Factors 2015", 2015, <https://moz.com/search-ranking-factors> (Última visita: 09/07/2017)
- [59] Google Webmaster Central Blog, "Using site speed in web search ranking", 9 Abril 2010, <https://webmasters.googleblog.com/2010/04/using-site-speed-in-web-search-ranking.html> (Última visita: 09/07/2017)
- [60] Jennifer Slegg, "Matt Cutts: Be Careful About How You Choose Your ccTLDs", Search Engine Watch, 1 Agosto 2013, <https://searchenginewatch.com/sew/how-to/2286477/matt-cutts-be-careful-about-how-you-choose-your-cclds> (Última visita: 09/07/2017)
- [61] Barry Schwartz, "Google's Matt Cutts: We Dropped The 100 Links Per Page Guideline But We May Take Action If It Is Too Spammy", Search Engine Land, 25 Noviembre 2015, <http://searchengineland.com/googles-matt-cutts-we-dropped-the-100-links-per-page-guideline-but-we-may-take-action-if-it-is-too-spammy-178197> (Última visita: 09/07/2017)
- [62] Statista, "Average number of search terms for online search queries in the United States as of February 2017", Febrero 2017 <https://www.statista.com/statistics/269740/number-of-search-terms-in-internet-research-in-the-us/> (Última visita: 09/07/2017)
- [63] Enge E., Spencer S., y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition, pp. 195-199.
- [64] Enge E., Spencer S., y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition, pp. 199-200.
- [65] Enge E., Spencer S., y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition, pp. 201.
- [66] Google AdWords Keyword Planner, <https://adwords.google.com/KeywordPlanner> (Última visita: 09/07/2017)
- [67] Bing Keyword Research, <https://www.bing.com/toolbox/keywords> (Última visita: 09/07/2017)
- [68] Google Trends, <https://trends.google.es/trends/> (Última visita: 09/07/2017)
- [69] Keyword Density Analyzer, Seobook, <http://tools.seobook.com/general/keyword-density/> (Última visita: 09/07/2017)
- [70] Enge E., Spencer. S, y Stricchiola J.C. (2015). "The Art of SEO", 3rd Edition, pp. 260-262.
- [71] Google Webmaster Central Blog, "Understanding web pages better", 23 Mayo 2014, <https://webmasters.googleblog.com/2014/05/understanding-web-pages-better.html> (Última visita: 09/07/2017)
- [72] XML-Sitemap Generator, <https://www.xml-sitemaps.com/> (Última visita: 09/07/2017)
- [73] Google Webmasters, "Mobile friendly websites", <https://developers.google.com/webmasters/mobile-sites/mobile-seo/?hl=es-419> (Última visita: 09/07/2017)
- [74] Google Webmasters, "Separate URLs", <https://developers.google.com/webmasters/mobile-sites/mobile-seo/separate-urls> (Última visita: 09/07/2017)
- [75] Google Webmasters, "Dynamic serving", <https://developers.google.com/webmasters/mobile-sites/mobile-seo/dynamic-serving> (Última visita: 09/07/2017)
- [76] Google Webmasters, "Responsive design", <https://developers.google.com/webmasters/mobile-sites/mobile-seo/responsive-design> (Última visita: 09/07/2017)
- [77] Google PageSpeed Insights, <https://developers.google.com/speed/pagespeed/insights/> (Última visita: 09/07/2017)
- [78] GTmetrix, <https://gtmetrix.com/> (Última visita: 09/07/2017)

- [79] Google, “Habilitar compresión”,
<https://developers.google.com/speed/docs/insights/EnableCompression> (Última visita: 09/07/2017)
- [80] Apache, “Apache Module mod_deflate”,
http://httpd.apache.org/docs/current/mod/mod_deflate.html (Última visita: 09/07/2017)
- [81] HTTP Compression Test, <http://www.whatsmyip.org/http-compression-test/> (Última visita: 09/07/2017)
- [82] Google, “Especificar caché navegador”,
<https://developers.google.com/speed/docs/insights/LeverageBrowserCaching> (Última visita: 09/07/2017)
- [83] w3.org, “Header field definitions”, <https://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html>
(Última visita: 09/07/2017)
- [84] Apache, “Apache Module mod_expires”,
http://httpd.apache.org/docs/current/mod/mod_expires.html (Última visita: 09/07/2017)
- [85] Google, “Minificar recursos”, <https://developers.google.com/speed/docs/insights/MinifyResources>
(Última visita: 09/07/2017)
- [86] Google, “Optimizar imágenes”,
<https://developers.google.com/speed/docs/insights/OptimizeImages> (Última visita: 09/07/2017)
- [87] Google, “Quitar el JavaScript que bloquea la visualización de contenido”,
<https://developers.google.com/speed/docs/insights/BlockingJS> (Última visita: 09/07/2017)
- [88] Google, “Optimizar la entrega de CSS”,
<https://developers.google.com/speed/docs/insights/OptimizeCSSDelivery> (Última visita: 09/07/2017)
- [89] Google, “Introduction to Structured Data”, <https://developers.google.com/search/docs/guides/intro-structured-data> (Última visita: 09/07/2017)
- [90] Schema.org, “BreadcrumbList”, <http://schema.org/BreadcrumbList> (Última visita: 09/07/2017)
- [91] Schema.org, “BreadcrumbList”, <http://schema.org/WebSite> (Última visita: 09/07/2017)
- [92] Schema.org, “EducationalOrganization”, <http://schema.org/EducationalOrganization> (Última visita: 09/07/2017)
- [93] Schema.org, “CreativeWork”, <http://schema.org/CreativeWork> (Última visita: 09/07/2017)
- [94] Schema.org, “NewsArticle”, <http://schema.org/NewsArticle> (Última visita: 09/07/2017)
- [95] Schema.org, “Article”, <http://schema.org/Article> (Última visita: 09/07/2017)
- [96] Schema.org, “ScholarlyArticle”, <http://schema.org/ScholarlyArticle> (Última visita: 09/07/2017)
- [97] Google Structured Data Testing Tool, <https://search.google.com/structured-data/testing-tool>
(Última visita: 09/07/2017)
- [98] Open Graph Protocol, <http://ogp.me/> (Última visita: 09/07/2017)
- [99] Twitter Developer Documentation, “Twitter Cards”, <https://dev.twitter.com/cards/overview> (Última visita: 09/07/2017)
- [100] Google Scholar, “Inclusion”, <https://scholar.google.com/intl/en/scholar/inclusion.html> (Última visita: 09/07/2017)
- [101] Adobe Acrobat, “Propiedades y metadatos del archivo PDF”,
http://help.adobe.com/es_ES/acrobat/using/WS58a04a822e3e50102bd615109794195ff-7c67.w.html
(Última visita: 09/07/2017)
- [102] Multiscale materials modeling group - UVa, “Molecular Implants”,
<https://www.ele.uva.es/~mmm/research/irradiation-of-materials/molecular-implants.html> (Última visita: 09/07/2017)

[103] Multiscale materials modeling group - UVa, "Classical molecular dynamics",
<https://www.ele.uva.es/~mmm/research/atomistic-modeling-techniques/classical-molecular-dynamics.html> (Última visita: 09/07/2017)

7. Anexos

Tabla A.1. *Posicionamiento previo*

Palabra clave	Google	Bing	Palabra clave	Google	Bing
Multiscale materials modeling	32	53	Atomistic modeling techniques	18	8
Projects	-	-	Ab initio	-	-
Publications	-	-	Binary collision approximation	-	-
Research	-	-	Classical molecular dynamics	-	-
Irradiation of materials	50	19	Kinetic monte carlo	-	-
Damage generation mechanisms	-	12(p)	Tight-binding	-	-
Description of irradiation damage	-	-	Technological processes	-	-
Molecular implants	55	-	Semiconductor doping	-	-
Influence of irradiation parameters	51	-	Defect engineering	-	-
Defects in semiconductors	48	76	Sputtering	-	-
Structural characterization	-	-	Nanodevices	-	-
Energetic characterization	52	-	Thermal treatments	-	-
Electronic properties	-	-	Amorphous films	-	-
Dynamical studies	-	-	Photoluminescence	-	-

Tabla A.2. *Lista palabras clave (1/3).*

Lista de palabras clave			
Palabra clave	Busq. mes	Palabra clave	Busq. mes
a-si	1k - 10k	charge trapping	100 - 1k
a-si:H	-	chemical bonds	10k - 100k
ab initio	10k - 100k	classical molecular	-
ab initio calculations	100 - 1k	classical molecular dynamics	10 - 100
ab initio simulations	-	cluster	100k - 1M
amorphization	-	clustering	-
amorphized	-	clustering in silicon	-
amorphizing	-	codiffusion	-
amorphous crystalline silicon interface	-	cold implants	-
amorphous films	10-100	collision approximation	-
amorphous layer	10-100	collision approximation kinetic	-
amorphous pockets	-	collision approximation kinetic monte carlo	-
amorphous semiconductors	100 - 1k	compactness	1k - 10k
amorphous silicon	10k - 100k	conduction band	1k - 10k
annealing	10k - 100k	crystal defects	1k - 10k
approximation kinetic monte carlo	-	crystal structure	10k - 100k
argon ions	100 - 1k	crystalline	10k - 100k
atomistic	1k - 10k	crystalline and amorphous	100 - 1k
atomistic modeling	10 - 100	crystalline silicon	1k - 10k
atomistic modeling techniques	-	damage	100k - 1M
atomistic origin	-	damage accumulation	-
atomistic simulation	100 - 1k	damage generation	-
b-doped	-	damage generation mechanisms	-
binary collision	10 - 100	defect	100k - 1M
binary collision approximation	10 - 100	defect clusters	-
bond defect	10 - 100	defect engineering	10 - 100
boron	10k - 100k	defect structures	-
boron activation	-	defects	100k - 1M
boron doped	100 - 1k	defects in semiconductors	10 - 100
boron interstitial clusters	-	defects in silicon	10 - 100
c-Si	100 - 1k	density	100k - 1M
characterization	10k - 100k	description irradiation damage	-

Tabla A.2. *Lista palabras clave (2/3).*

Lista de palabras clave			
Palabra clave	Busq. mes	Palabra clave	Busq. mes
desorption	1k - 10k	ion	100k - 1M
device performance	100 - 1k	ion beam	100 - 1k
diffusion	100k - 1M	ion implantation	1k - 10k
dissolution	10k - 100k	ion implantation technologies	-
distribution of defects	-	irradiation	10k - 100k
dopant	1k - 10k	irradiation damage	10 - 100
dopant diffusion	10 - 100	irradiation defect clusters	-
dopant incorporation	-	irradiation defects	-
dynamical studies	-	irradiation induced defects	-
dynamics binary collision	-	irradiation of materials	-
electronic properties	100 - 1k	irradiation parameters	-
energetic characterization	-	junctions	10k - 100k
energetic particles	10 - 100	kinetic monte carlo	100 - 1k
energy	100k - 1M	laser processes	100 - 1k
energy transfers	10k - 100k	locally strained	-
experimental technique	100 - 1k	locally strained regions	-
extended defects	-	materials	100k - 1M
extended defects structures	-	materials modeling	100 - 1k
finfet	10k - 100k	materials modeling group	-
fluorine	10k - 100k	modeling	100k - 1M
germanium	10k - 100k	modeling of defects	-
germanium crystallization	-	modeling research group	-
heterojunctions	1k - 10k	modeling techniques	1k - 10k
hydrogenated amorphous silicon	100 - 1k	molecular	10k - 100k
implantation	10k - 100k	molecular beam	100 - 1k
implants	100k - 1M	molecular dynamics	1k - 10k
improved bca	-	molecular dynamics simulations	100 - 1k
improved bca model	-	molecular implants	-
influence irradiation parameters	-	multiscale	1k - 10k
influence of irradiation	-	multiscale materials	10 - 100
interface structure	10 - 100	multiscale materials modeling	10 - 100
interstitial	10k - 100k	multiscale schemes	-
interstitial defects	100 - 1k	multiscale simulation scheme	-

Tabla A.2. *Lista palabras clave (3/3).*

Lista de palabras clave			
Palabra clave	Busq. mes	Palabra clave	Busq. mes
nanodevices	100 - 1k	semiconductor doping	100 - 1k
nanometric multigate	-	semiconductors	10k - 100k
nanowires	1k - 10k	silicon	100k - 1M
nucleation	10k - 100k	silicon devices	100 - 1k
octadecaborane	10 - 100	simulation	1M - 10M
partial amorphization	-	sputtering	10k - 100k
phase transition	1k - 10k	statistical analysis	10k - 100k
phosphorus	100k - 1M	strained regions	-
photoluminescence	1k - 10k	structural characterization	10 - 100
photoluminescence lines	-	submicrosecond laser processes	-
pocket	100k - 1M	tcad	10k - 100k
preamorphized	-	techniques	100k - 1M
preamorphized silicon	-	technological processes	100 - 1k
predictive modeling	1k - 10k	technology computer assisted design	-
publications	10k - 100k	tem	100k - 1M
radiation	100k - 1M	thermal	10k - 100k
radiation damage	100 - 1k	thermal annealing	100 - 1k
recrystallization	10k - 100k	thermal conductivity	10k - 100
recrystallization processes	-	thermal treatments	100 - 1k
research	100k - 1M	tight binding	100 - 1k
residual damage	100 - 1k	tight binding simulations	-
ripening	1k - 10k	transmission electron microscopy	-
schrödinger equation	10k - 100k	ultrashallow	-
self interstitial	100 - 1k	uphill diffusion	100 - 1k
self interstitial clustering	-	v-shaped	10k - 100k

Tabla A.3. Optimización de las imágenes.

Imagen	Original	Óptimo	Ganancia
intro-mmm.png	297 KB	81 KB	72,73%
intro-defects-semiconductors.png	465 KB	118 KB	74,62%
intro-irradiation-materials.png	904 KB	75 KB	91,70%
intro-amorphous-films.png	404 KB	109 KB	73,02%
intro-nanodevices.png	694 KB	161 KB	76,80%
intro-semiconductor-doping.png	164 KB	69 KB	57,93%
logo_mmm.png	6 KB	4 KB	33,33%
logo_uva.png	3 KB	2 KB	33,33%
variables-network.png	20 KB	7 KB	65,00%
UIC_JCyL.png	193 KB	15 KB	92,23%
logoiit.png	8 KB	5 KB	37,50%
res-atomistic-modeling-techniques.png	314 KB	101 KB	67,83%
res-amt-ab-initio.png	475 KB	95 KB	80,00%
res-amt-tight-binding.png	760 KB	151 KB	80,13%
res-amt-classical-molecular-dynamics.gif	1,6 MB	943 KB	39,94%
res-amt-binary-collision-approximation.png	797 KB	197 KB	75,28%
res-amt-kinetic-monte-carlo.png	201 KB	100 KB	50,25%
res-irr-damage-mechanisms.png	101 KB	24 KB	76,24%
res-irr-typology.png	885 KB	320 KB	63,84%
res-irr-molecular-implants.png	449 KB	300 KB	33,18%
res-irr-parameters.png	203 KB	187 KB	7,88%
res-def-structural.png	589 KB	93 KB	84,21%
res-def-structural-FinFET.png	305 KB	271 KB	11,15%
res-def-energetic.png	122 KB	104 KB	14,75%
res-def-energetic-defects.png	882 KB	104 KB	88,21%
res-def-electronic.png	260 KB	223 KB	14,23%
res-def-dynamical-studies.gif	2,5 MB	1,7 MB	31,08%
res-proc-doping-sheet-resistance.png	23 KB	14 KB	39,13%
res-proc-doping-uphill.png	57 KB	23 KB	59,65%
res-proc-defect-engineering.png	243 KB	31 KB	87,24%
res-proc-sputtering.png	261 KB	83 KB	68,20%
res-proc-nanodevices-FinFET-schema.png	66 KB	15 KB	77,27%
res-proc-nanodevices-FinFET-doping.png	431 KB	41 KB	90,49%
res-proc-nanodevices-Finfet110.png	352 KB	113 KB	67,90%
res-proc-nanodevices-Finfet100.png	231 KB	73 KB	68,40%
res-proc-thermal-treatments.png	336 KB	36 KB	89,29%
res-proc-amorphous-films.png	880 KB	109 KB	87,61%
	16412 KB	6093 KB	62,87%

Tabla A.4. Evolución mensual del posicionamiento (1/2).

	Marzo '16				Abril '16				Mayo '16				Junio '16			
	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM
1-3	2	1	-	-	1	0	4	2	1	1	2	3	1	2	2	3
4-10	2	3	-	-	7	3	7	9	8	3	6	8	7	3	6	8
11-20	1	1	-	-	6	1	3	8	6	1	4	6	5	2	3	3
21-50	7	3	-	-	11	3	6	12	9	6	4	10	7	6	4	9
51-120	11	4	-	-	14	6	7	13	19	8	9	14	19	6	6	15
Total	23	13	-	-	39	13	28	43	42	18	25	41	38	18	20	37
Papers	9	6	-	-	9	4	12	3	12	5	11	10	10	3	12	9
Publi.	0	0	-	-	6	0	0	8	5	5	0	6	2	6	0	3
	Julio '16				Agosto '16				Septiembre '16				Octubre '16			
	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM
1-3	4	3	2	6	5	4	5	5	5	4	2	5	4	5	0	5
4-10	6	5	5	8	6	6	4	6	7	6	5	7	10	5	6	10
11-20	6	3	3	4	9	3	4	8	8	4	3	7	6	5	3	6
21-50	12	12	2	13	10	13	3	12	12	11	3	11	13	12	5	12
51-120	17	7	3	12	17	10	3	14	16	10	3	15	15	10	4	13
Total	44	29	16	42	47	35	19	45	47	35	17	45	48	35	17	46
Papers	11	3	10	10	12	5	10	10	13	7	12	12	13	6	13	12
Publi.	1	7	0	1	1	11	3	1	2	7	1	2	5	9	1	5

*G=Google, B=Bing, Y=Yandex, GM=Google Mobile

Tabla A.4. Evolución mensual del posicionamiento (2/2).

	Noviembre '16				Diciembre '16				Enero '17				Febrero '17			
	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM
1-3	6	4	1	8	6	3	1	7	5	4	2	5	2	4	1	3
4-10	7	6	8	7	7	6	8	7	6	5	5	8	9	3	6	9
11-20	10	3	3	9	11	4	2	10	12	3	1	12	12	3	2	13
21-50	15	12	4	16	14	12	5	15	13	13	6	15	10	8	7	12
51-120	14	10	6	10	13	10	4	11	17	6	3	13	18	5	5	14
Total	52	36	22	50	51	34	21	50	52	32	16	52	51	24	21	50
Papers	13	7	10	11	9	9	13	10	8	7	10	8	6	6	15	7
Publi.	7	9	4	7	8	6	1	7	10	5	1	10	11	5	1	11
	Marzo '17				Abril '17				Mayo '17				Junio '17			
	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM	G	B	Y	GM
1-3	2	3	3	3	4	3	3	5	3	2	3	5	5	5	2	5
4-10	8	3	6	7	8	3	6	8	6	5	3	6	9	5	4	9
11-20	11	4	2	11	9	4	2	9	10	4	1	10	7	3	1	7
21-50	8	8	8	10	8	9	6	8	7	8	8	9	8	7	4	9
51-120	20	8	9	18	21	10	5	19	15	8	6	14	16	8	4	15
Total	50	26	27	50	49	29	21	49	41	28	21	44	46	28	14	45
Papers	5	5	15	6	6	8	13	6	2	7	19	3	3	6	13	3
Publi.	13	7	2	13	11	4	0	12	14	6	0	16	16	6	0	15

*G=Google, B=Bing, Y=Yandex, GM=Google Mobile

Tabla A.5. Evolución mensual del posicionamiento (España)

	Julio '16		Agosto '16		Sep. '16		Octubre '16		Nov. '16		Dic. '16	
	G	B	G	B	G	B	G	B	G	B	G	B
1-3	5	28	7	32	7	27	8	32	9	32	9	33
4-10	10	13	10	14	11	15	9	19	10	20	10	17
11-20	2	8	3	7	3	9	5	7	5	8	5	9
21-50	14	6	13	4	12	4	11	3	11	6	11	4
51-120	17	2	16	3	15	5	14	3	11	3	11	3
Total	48	57	48	60	47	60	46	65	46	69	46	66
Papers	15	5	13	4	15	8	14	7	13	8	13	8
Publi.	2	21	2	28	2	24	4	29	5	27	5	27
	Enero '17		Febrero '17		Marzo '17		Abril '17		Mayo '17		Junio '17	
	G	B	G	B	G	B	G	B	G	B	G	B
1-3	7	28	6	17	7	24	9	30	8	31	10	40
4-10	10	16	11	16	12	16	10	18	11	17	10	14
11-20	7	4	7	4	7	6	7	7	5	5	5	5
21-50	14	3	10	3	13	4	8	3	7	2	8	2
51-120	11	1	12	0	13	2	13	4	13	2	13	4
Total	49	51	47	40	53	51	47	63	44	57	46	65
Papers	11	7	8	7	7	5	6	5	5	4	5	4
Publi.	7	15	9	8	13	19	10	22	12	24	13	27

*G=Google, B=Bing

Contenido CD

- Memoria.pdf
- Seguimiento/
 - Herramienta de visualización: “SEO – Visualización.xlsm”
 - Logs de seguimiento/
 - Marzo 2016
 - Abril 2016
 - Mayo 2016
 - Junio 2016
 - Julio 2016
 - Agosto 2016
 - Septiembre 2016
 - Octubre 2016
 - Noviembre 2016
 - Diciembre 2016
 - Enero 2017
 - Febrero 2017
 - Marzo 2017
 - Abril 2017
 - Mayo 2017
 - Junio 2017
 - Estado indexación
 - Seguimiento títulos publicaciones

La herramienta de visualización “SEO – Visualización.xlsm” es un archivo Excel que nos permite hacer un análisis visual de los datos obtenidos en el proceso de seguimiento. Como podemos ver en la Figura A.1, la herramienta consta de un menú desplegable en el que podemos seleccionar la palabra clave cuyos datos deseemos visualizar. Una vez seleccionada la visualización de los datos es inmediata. En concreto la herramienta nos ofrece la posibilidad de visualizar los datos completos de todo el proceso de seguimiento o bien visualizar los datos de cada uno de los meses navegando a través del menú.



Figura A.1. Herramienta de visualización.