

Búsqueda y Evaluación de Información para Periodistas y Comunicadores



Lluís Codina, Rafael Pedraza, Cristòfol Rovira, Carlos Lopezosa
Àrea de Formatos Digitales e Interactivos
[Grupo DigiDoc](#)

Departamento y Facultad de Comunicación
Universitat Pompeu Fabra

Búsqueda y evaluación de información para periodistas y comunicadores

Por **Lluís Codina, Rafael Pedraza, Cristòfol Rovira, Carlos Lopezosa**

Área de Formatos Digitales e Interactivos

[Grupo DigiDoc](#)

Departamento y Facultad de Comunicación

Universitat Pompeu Fabra

Edición: enero 2020



Esta obra está bajo una [Licencia Creative Commons Atribución 4.0 Internacional](#).

0. Introducción

Periodistas y comunicadores necesitan conocer lo mejor posible la naturaleza del medio donde desarrollan buena parte de su actividad. Este medio es Internet, la Web o el ecosistema digital, como queramos llamarle.

En el caso de la comunicación social, periodistas y profesionales de la comunicación audiovisual comparten la doble necesidad de saber utilizar los sistemas de búsqueda en la web y de evaluar la calidad de los páginas o sitios encontrados en sus actividades de búsqueda.

Para los profesionales intensivos en información, como lo son periodistas y comunicadores del audiovisual, es imperativo saber resolver sus necesidades de información de forma profesional y eficiente.

Además, buena parte de estos profesionales van a tener la necesidad de llevar a cabo algún tipo de emprendimiento en el mundo digital, lo que conlleva también la de conocer al menos los fundamentos del posicionamiento web, también denominado SEO.

En esta obra, llevada a cabo con objetivos claramente docentes, se introduce a los estudiantes de Periodismo y de Comunicación Audiovisual en los fundamentos de la búsqueda en Internet, se presenta una primera introducción al posicionamiento web y por último se ofrecen nociones sobre evaluación de sitios web.

Los diferentes apartados van acompañados de Actividades cuyo buen desempeño proporcionarán al estudiante una base sólida de aprendizaje de las tres materias de las que este documento pretende ser una introducción para posteriores desarrollos.

1. Búsqueda

Antes de introducirnos en el uso efectivo de los buscadores debemos considerar cómo funcionan y cuál es la anatomía de un buscador. El objetivo es tener un grado de comprensión de su naturaleza, aunque sea a mil metros de altura, que nos ayude a utilizarlos mejor. En realidad, es difícil imaginar el papel de los medios actualmente sin su relación con los buscadores. Conocerlos mejor nos ayudará no solamente a ser más eficientes encontrando información para nuestro trabajo como periodistas o comunicadores, sino también a que nuestras producciones periodísticas tengan mayor visibilidad en la web.

La cuestión es que los motores de búsqueda son los más importantes sistemas de información de nuestro tiempo. Forman parte del paisaje de Internet de una manera que parece definitiva. Nadie imagina hoy la posibilidad de explotar la inmensa riqueza de la Web sin el auxilio de los motores de búsqueda cuyas funciones no han dejado ni de aumentar ni de experimentar cambios desde su temprana aparición en los años 90.

Sistemas de información documental

Una forma que puede resultar útil para concebir a los buscadores consiste en considerarlos como una clase de sistema de información documental. El calificativo documental se refiere a que buscan y analizan informan registrada en documentos, esto es, en páginas web, por ejemplo.

Los sistemas de información documentales se pueden representar con el diagrama (Ilustración 1) que mostramos a continuación:

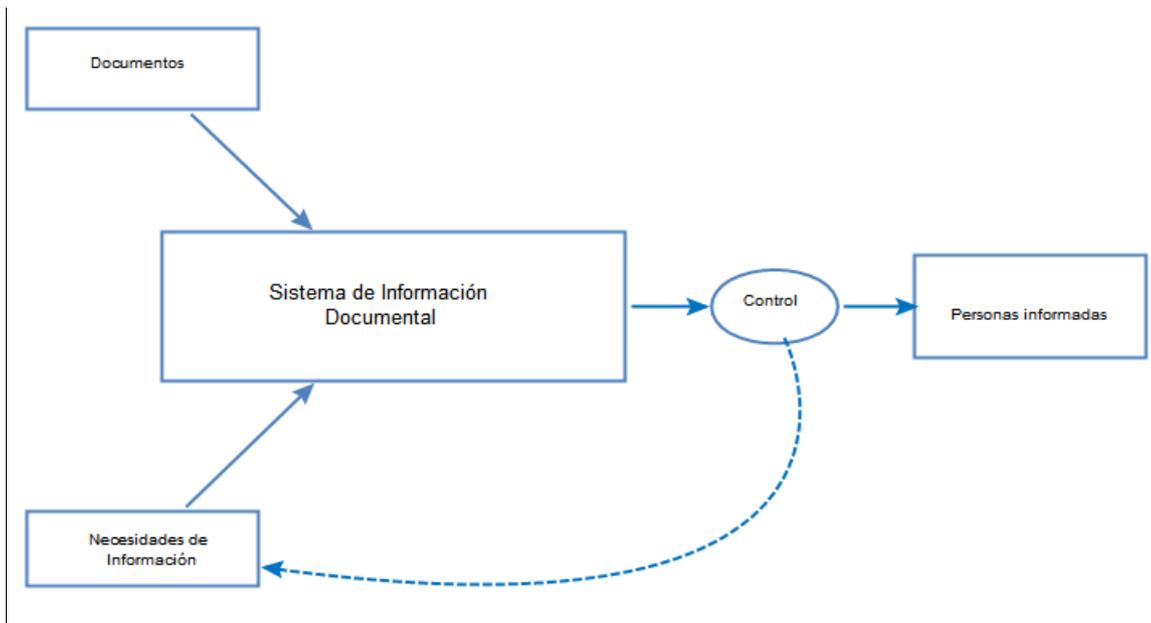


Ilustración 1: Diagrama general de un Sistema de Información documental

Podemos ver que un sistema documental (tal como un buscador) acepta dos clases de entradas: **Documentos** y **Necesidades de Información**. Los documentos en nuestro caso son páginas web (pero no únicamente, ya que los servidores web incluyen también documentos en pdf u otros formatos) y las necesidades de información son las palabras clave o preguntas de los usuarios.

En un buscador, la salida del sistema es una lista de enlaces a sitios web que, presumiblemente, contienen información capaz de satisfacer las necesidades de información del usuario, de aquí que la salida se rotule con el (optimista) texto de **Personas informadas**. En nuestro caso, **Periodistas Informados**.

Naturalmente, si la primera consulta no da buenos resultados, lo lógico es que el usuario (el periodista, p.e.) haga algún intento más. Esto queda representado por la elipse rotulada con la palabra **Control**. Es decir, se supone que el usuario examina la relevancia de los resultados y si no le satisfacen, reitera la pregunta usando otros términos o parametrizando la consulta mediante la búsqueda avanzada.

Estructura de un motor de búsqueda

Un buscador consiste en un conjunto de programas que trabajan en cooperación. Estos programas se pueden distribuir en dos grupos principales:

- *Sistema de exploración, también llamado Rastreador (spyder o crawler)*
- *Sistema de recuperación de información (information retrieval system)*

Cada uno de estos sistemas se compone de otros subsistemas (Ilustración 2):

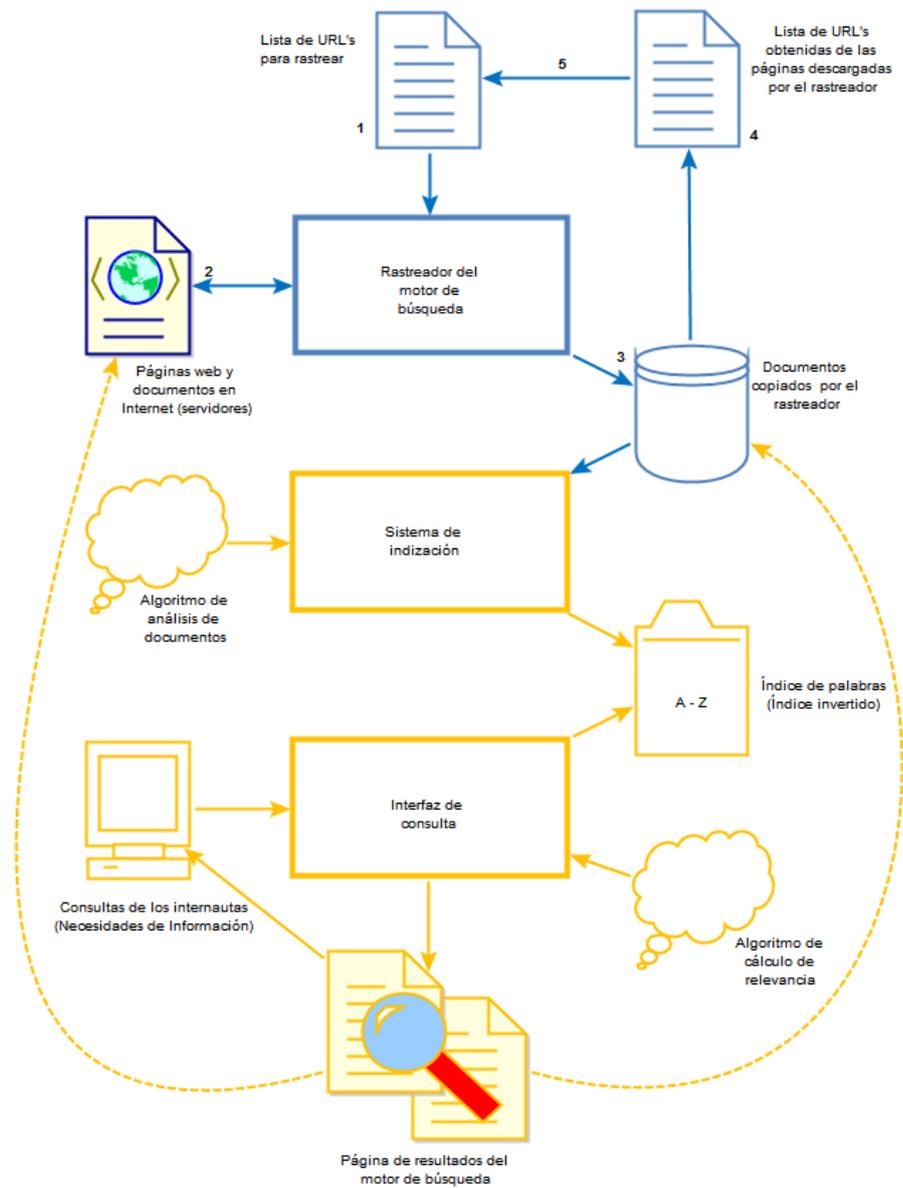


Ilustración 2: Estructura y funciones de un buscador (search engine)

Rastreador

El rastreador, o *sistema de exploración* tiene la misión de descubrir y descargar páginas web y otros documentos de Internet. Para ello, parte de una lista inicial de URL o direcciones de páginas web (1). A partir de ella accede a documentos (2) publicados en la web. Descarga esos documentos, en su mayoría páginas web, pero también documentos ofimáticos, presentaciones, archivos de imágenes, etc. en el almacén o repositorio de documentos del Buscador (3). A partir de las páginas descargadas, el rastreador obtiene nuevas URL (4) que añade a su lista de URL pendientes de analizar (5). Cuando producimos una página web nueva, nuestra esperanza es que los Buscador la detecten y la indexen. Pero no tenemos porqué esperar a que esto suceda. Podemos tomar la iniciativa y avisar a los Buscador, y en particular a Google de que tal página existe y conseguir así que la incluya en sus índices.

Actividad 1

Vaya a la página <https://www.google.com/search/howsearchworks/crawling-indexing/>. Revise la información que Google proporciona sobre cómo funciona el rastreo. Lea el artículo (ver el vídeo es opcional) y haga un resumen del contenido de entre 150 y 200 palabras (si es más extenso, no es problema) sobre lo aprendido, acompañado de una captura de pantalla de la página y la URL de la fuente.

Sistema de recuperación de información

El sistema de recuperación de información tiene a su vez dos componentes:

- Sistema de indexación
- Sistema de consulta

El primero de ellos se ocupa del análisis de los documentos descargados de la web y de la creación de los índices que permiten luego las consultas al motor; mientras que el segundo constituye la interfaz visible del BUSCADOR, precisamente la parte con la que interactúan los usuarios.

Sistema de indexación

Si el buscador puede responder preguntas en el asombrosamente breve tiempo al que nos ha acostumbrado tanto Google como Bing o Qwant (típicamente, fracciones de segundo), es porque los buscadores no exploran la web en tiempo real.

Un sistema que pretendiera explorar de forma secuencial los documentos de la web para ofrecer respuestas no tendría el más mínimo sentido. Incluso con la potencia de cálculo de los ordenadores actuales, podrían transcurrir meses entre la pregunta y la respuesta. Un estudiante que hiciera su pregunta en el primer trimestre de la carrera, cuando obtuviese la respuesta ya estaría en segundo curso (eso si no suspendió la asignatura esperando la respuesta del buscador). En lugar de lo anterior, lo que hacen los buscadores es consultar sus

índices internos.

Actividad 2: Haga diversas consultas (dos o tres es suficiente) con palabras relativamente habituales (como televisión, ordenadores, música, etc.) en Google o en Bing y compare los tiempos de respuestas y el número aproximado de documentos encontrados (ambas cosas indicadas en una barra situada justo sobre la lista de resultados). Luego haga dos o tres preguntas con términos más especializados (como "conductas adictivas", "derecho al olvido", "refugiados guerra de Siria", etc.) Compruebe si, con diferentes preguntas, los tiempos de respuesta son muy distintos o, en general, los tiempos son similares. Cabe esperar que no haya diferencias apreciables. El motivo se discutirá más adelante. Haga capturas de pantalla comentadas.

Índices directos

Los buscadores utilizan al menos dos clases de índices, denominados respectivamente índices directos e índices invertidos. Las consultas las resuelven los *índices invertidos*, mientras que los *índices directos*, que son los que vamos a ver ahora, se utilizan como elementos de gestión y control internos.

En un índice directo tenemos la lista de documentos (o de registros) en un orden cronológico (el documento más antiguo primero, por ejemplo) o numérico (del documento número 1 hasta el último). La siguiente tabla ilustra esta clase de índices mediante un ejemplo de índice directo:

Núm. Documento	Contenido
00001	Título: Manual de lenguajes documentales Autor: Blanca Gil
00002	Título: Diseño y construcción de sitios con Dreamweaver 8 Autor: Marc Campbell
...	...
12345	Título: Aprender a razonar Autor: Fina Pizarro

Representación de un índice directo

Como podemos ver, con este índice, para saber si hay un documento con las palabras "aprender" y "razonar", tendríamos que recorrer decenas de miles de entradas del índice (concretamente 12345). Lo peor es que, si el índice completo tuviera, por ejemplo, cien mil entradas, habría que recorrer las cien mil entradas del índice para saber si hay más de un

documento que cumpla la condición anterior.

Como es fácil suponer, esta clase de índices no serviría para la búsqueda en Internet. Por tanto, un buscador necesita complementar este índice con un índice invertido, que es el que se utiliza realmente para responder a las consultas (mientras que el índice directo se utiliza para aspectos de gestión y administración internos).

Índices invertidos

La estructura de un índice invertido es exactamente la inversa de la anterior (de aquí el nombre), a saber: consiste en una lista ordenada de todas y cada una de las palabras que aparecen en los distintos documentos (páginas web en el caso de Internet) asociadas a la lista concreta de los documentos (páginas) en los que aparecen. La estructura típica de un índice invertido, por tanto, es la siguiente:

<i>Término único</i>	<i>Frecuencia</i>	<i>Ubicación</i>
Aprender	110	(12345, 01, 01) ...
...
Blanca	233	(00001, 02, 01) ...
...
Campbell	6	(00002, 02, 02) ...
...
Razonar	214	(12345, 01, 03) ...
...

Representación de un fichero invertido

La explicación:

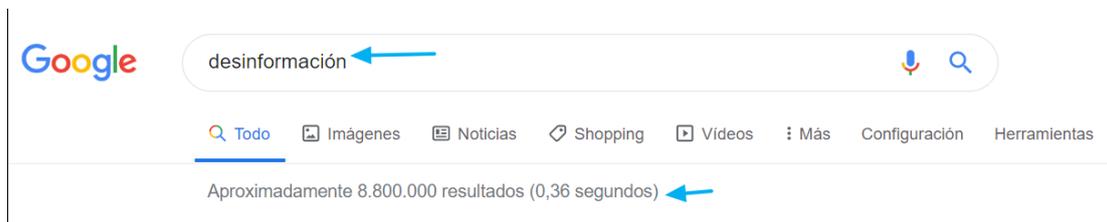
- En ***Término único*** aparecen las distintas palabras de los documentos, en base a una sola fila por cada palabra (aunque en el conjunto de los documentos la palabra aparezca muchas veces).
- En la columna ***Frecuencia*** tenemos el número total de veces que aparece cada término.
- Por último, en ***Ubicación*** tenemos una clave donde aparece el número de documento,

la zona o campo donde aparece la palabra y el orden de la palabra. Esto se denomina **vector**, y habrá uno por cada ocurrencia de palabra.

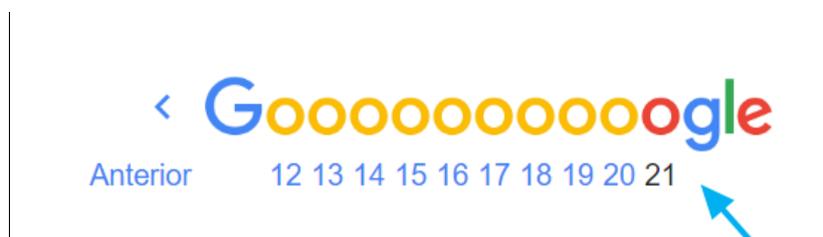
Veamos esto último con más detalle. Tomemos el término “Aprender”. Tenemos que la columna de **Frecuencia** señala: “110”. Esto significa que “Aprender” aparece 110 veces en el conjunto de los documentos, y que habrá por tanto 110 vectores distintos en la columna **Ubicación**. Nosotros solamente hemos representado uno de ellos para simplificar. Tomemos ahora este mismo vector (34567, 01, 01) y veámoslo con detalle. El primer número, 34567, es el que identifica el documento, el segundo, 01, identifica el campo o zona del documento donde aparece la palabra, en este caso, **Título**. El tercer número, 01, dice que es la **primera** palabra del campo Título.

Otros índices auxiliares ayudan al Buscador a recorrer la lista de términos únicos con un pequeño número de operaciones de comparación. Lo mejor de todo es que esta clase de índices hace que el tiempo de respuesta sea virtualmente independiente del número de términos que aparecen en la lista.

El problema principal con estos índices es la cantidad de recursos de hardware que requieren, particularmente, si se debe dar servicio a miles (o millones) de usuarios simultáneamente. Por ejemplo, para la búsqueda [desinformación], Google afirma encontrar casi 9 millones de resultados:



Sin embargo, cuando intentamos avanzar por los resultados, al llegar al resultado número 200 aproximadamente ya no podemos seguir viendo más:



Actividad 3

Google utiliza alguna versión del tipo de índice invertido que hemos mostrado. Para ver que estos índices son eficaces, pero no permiten hacer milagros, haga una consulta en Google por un término muy frecuente como hizo antes (p.e. televisión) y compruebe hasta dónde puede llegar (cada página contiene unos 10 resultados). En el ejemplo anterior no hemos podido pasar de la página 21 (por tanto, unos 200 resultados). El motivo no está documentado oficialmente, pero es fácil deducir que la red de ordenadores de Google tiene una cantidad de memoria limitada que puede asignar a cada usuario y por tanto no puede

cargar toda la información de los resultados. Haga capturas de pantalla comentadas de las pruebas que ha hecho hasta llegar al final de la navegación por las páginas de resultados.

Interfaz de consulta

Los componentes son los siguientes:

- En primer lugar, hay un formulario simple o **avanzado** de consulta (como el de la captura que sigue) que sirve para enviar la pregunta a la parte del sistema que compara las palabras clave de la pregunta con el índice del buscador:

The image shows the Google Advanced Search interface. At the top left is the Google logo, and at the top right is a grid icon and a profile picture. Below the logo, the text "Búsqueda avanzada" is displayed in red. The main section is titled "Buscar páginas con..." and contains five rows of search criteria, each with a text input field and a corresponding instruction:

- todas estas palabras: Escribe las palabras importantes: terrier ratonero tricolor
- esta palabra o frase exactas: Escribe las palabras exactas entre comillas: "terrier ratonero"
- cualquiera de estas palabras: Escribe OR entre todas las palabras que quieras: miniatura OR estándar
- ninguna de estas palabras: Añade un signo menos delante de las palabras que no quieras que aparezcan: -roedor, -"Jack Russell"
- números desde el: hasta Escribe dos puntos seguidos entre los números y añade una unidad de medida: 10..35 kg, 300..500 euros, 2010..2011

Below this section, there is a heading "A continuación, limitar los resultados por..." followed by several filter options, each with a dropdown menu and a description:

- idioma: cualquier idioma (dropdown) Busca páginas en el idioma que selecciones.
- región: cualquier región (dropdown) Busca páginas publicadas en una región determinada.
- última actualización: en cualquier momento (dropdown) Busca páginas actualizadas en el periodo de tiempo especificado.
- sitio o dominio: Busca un sitio (como wikipedia.org) o limita los resultados a un dominio como, por ejemplo, .edu, .org o .gov.
- los términos que aparecen: en cualquier lugar de la página (dropdown) Busca términos en toda la página, en el título de la página o en la dirección web, o enlaces a la página que estás buscando.
- Búsqueda Segura: Ver resultados más relevantes (dropdown) Indica a [Búsqueda Segura](#) si quieres que filtre contenido sexualmente explícito.
- tipo de archivo: cualquier formato (dropdown) Busca páginas en el formato que prefieras.
- derechos de uso: sin filtrar por licencia (dropdown) Busca páginas que puedas utilizar libremente.

At the bottom right of the filter section is a blue button labeled "Búsqueda avanzada". Below the filters, there is a section titled "También puedes..." with four links: "Buscar páginas similares a una URL", "Buscar páginas visitadas", "Usar operadores en el cuadro de búsqueda", and "Personalizar configuración de búsqueda". At the very bottom of the page, the word "España" is displayed.

- En segundo lugar, el buscador debe presentar los resultados de la forma más clara y eficiente posible a través de la *página de resultados del motor de búsqueda*, o **SERP**, (por *search engine results page*).
- A partir de aquí, los resultados deben mostrarse en algún **orden significativo**, y de entrada podemos descartar el orden alfabético o el cronológico, dada su escasa utilidad en un contexto tan heterogéneo y con la inmensa cantidad de documentos que hay en Internet. Aquí interviene el denominado algoritmo de cálculo de relevancia, al que dedicaremos el resto de este apartado.
- Por último, la SERP (o página de resultados del buscador) debe estructurar y presentar una gran cantidad de información en un espacio muy reducido. En la captura siguiente, vemos que para una búsqueda sobre los Premios Gaudí del 2020, Google presenta una SERP con hasta 10 secciones con informaciones muy diferentes que debemos saber reconocer (ver página siguiente):

Aproximadamente 2.470.000 resultados (0,48 segundos)

Noticias destacadas



Premios Gaudí 2020: Los protagonistas de la gran noche del cine catalán

Lecturas
Hace 10 horas



Belén Funes y Marques-Marcet triunfan en unos premios Gaudí muy...

20Minutos
Hace 7 horas



'La hija de un ladrón' y 'Els dies que vindran' acaparan los Premis Gaudí

El Periódico
Hace 5 horas

[→ Más sobre premios gaudi 2020](#)

XII Premis Gaudí
Ceremonia de premiación

Fecha: dom., 19 ene. • 9:50 p. m. CET

Presentadora: Anna Moliner

Nominados y ganadores

<p>Premio Gaudi a La Mejor Actr...</p>  <p>Laia Marull La inocencia</p> <p>Ganadora/Ganador</p>	<p>Premio Gaudi a</p>  <p>Marçal Cebrián La hija de un ladrón</p> <p>Ganadora/Ganador</p>
<p>Nora Navas</p>  <p>Dolor y gloria</p>	<p>Belén Funes</p>  <p>La hija de un ladrón</p> <p>Ganadora/Ganador</p>
<p>Julieta Serrano</p>  <p>Dolor y gloria</p>	<p>Laia Soler</p>  <p>La inocencia</p>
<p>Aina Clotet</p>  <p>7 Reasons to Run Away (fr...</p>	<p>David Desola</p>  <p>El hoyo</p>

Premis Gaudí 2020: Ganan La hija de un ladrón y Els dies ...
hace 6 horas - El cine catalán puede vivir ahogado por la precariedad económica, como ha denunciado hasta la extenuación la presidenta de la Acadèmia, ...
1 person - [View similar links](#)

Premios Gaudí 2020: Los protagonistas de la gran noche del ...
hace 11 horas - Los Premios Gaudí 2020 ya se han entregado. Los famosos galardones del cine catalán, que conceden la Acadèmia del Cinema Català, ...

Premis Gaudí 2020: Sigue la entrega de premios del cine ...
hace 16 horas - Sigue en directo la gala de los XII Premis Gaudí de este 2020, que se celebra este domingo 19 de enero en el Auditori del Fòrum CCIB.

Tendencia actual Ver otros 1



Antoni Gaudí



Els dies que vindran



Carlos Marques



Belén Funes

[Regístrate como responsable de este panel de información](#)

Videos



Així han arrencat els Premis Gaudí 2020 - Premis Gaudí

CCMA - hace 15 horas



Premis Gaudí 2020: Horario y dónde ver la gala del cine catalán

Acadèmia del Cinema La Vanguardia - hace 1 día



XII Gala dels Premis Gaudí. - Premis Gaudí

CCMA - hace 16 horas

Premios Gaudí 2020: 'Els dies que vindran' de Marqués ...
hace 13 horas - Premis Gaudí 2020: 'Els dies que vindran' de Marqués-Marcet gana unos XII Premis Gaudí muy repartidos. También ha sido protagonista el ...

XII Premis Gaudí - Acadèmia del Cinema Català
Els XII Premis Gaudí se celebraran el 19 de gener de 2020 La dotzena edició de la gran nit del cinema català s'avança una setmana i torna a l'Auditori Fòrum ...

Premis Gaudí 2020 - Catalan Films
Premis Gaudí 2020. 19/01/2020 - 19/01/2020. Lugar: Barcelona , Catalunya; Tipos: Animación, Documentales, Ficció; Actividades: Todas; Ámbito: ...

Premis Gaudí 2020 - CCMA
Premis Gaudí 2020 Viu la festa del cinema català a TV3 i Catalunya Ràdio: ... Anna Moliner: "La gala dels Premis Gaudí tindrà música, ironia i emoció".

Belén Funes y Marques-Marcet triunfan en unos premios ...
hace 8 horas - El director Carlos Marqués-Marcet recibe el premio Gaudí 2020 a la mejor película por 'Els dies que vindran'. MARTA PÉREZ / EFE.

Búsquedas relacionadas con premios gaudi 2020

premis gaudi 2020 nominats	nominats gaudi
premis gaudi 2019	premis gaudi 2018
nominacions gaudi 2020	xi premis gaudi
premis gaudi 2019 nominats	inscripcions premis gaudi

Goooooooooooooo**gle** >
1 2 3 4 5 6 7 8 9 10 [Siguiente](#)

De la captura precedente de la SERP de Google para la consulta sobre los Premios Gaudí 2020 podemos destacar los siguientes componentes:

1. Opciones para ir a índices especializados: noticias, imágenes, vídeos, etc.
2. Carrusel con noticias destacadas, tomadas de contenidos de medios de comunicación. Es necesario que estos medios lleven a cabo un buen SEO periodístico para figurar aquí, lo que les proporciona mucha visibilidad y mucho más tráfico que a otros medios.
3. Datos factuales de consulta rápida de los Premios Gaudí del 2020. Puede servir para responder la intención de la consulta si era obtener estos datos. Evita tener que entrar en páginas web.
4. Lista de resultados procedentes de medios de comunicación. Para estar en la primera página de resultados, estos sitios han tenido que hacer también muy buen SEO periodístico. Aparecer en estas posiciones garantiza mucho más tráfico a estos medios.
5. Respuestas directas con datos de nominados y ganadores. Como en el anterior caso, evitan tener que ir a la página web si buscamos datos concretos y los encontramos aquí.
6. Carrusel de vídeos (normalmente, proceden de YouTube y algunos pueden ser de medios de comunicación).
7. Temas que Google identifica como tendencias. Puede resultar útil en determinados temas.
8. Resto de páginas web que Google considera que son útiles para resolver la intención de búsqueda. Pueden ser (o no) de medios de comunicación.
9. Ideas para búsquedas relacionadas. Puede ser muy útil si seguimos necesitando información sobre el tema, pero dándole otros enfoques.
10. Navegación entre las páginas resultados si necesitamos explorar más resultados. Puede ser muy útil en búsquedas exhaustivas.

Actividad 4

Haga una búsqueda, tanto en Google como en Bing por algún tema de su interés, pero que tenga una dimensión cultural, política o social importante. Por ejemplo, puede buscar por nombres de premios en el ámbito audiovisual, como los premios Oscar, o los Goya o puede repetir la búsqueda sobre los premios Gaudí pero cambiando el año y poniendo los del 2019, por ejemplo (también puede utilizar nombres de artistas o de políticos muy conocidos). Examine la página de resultados o SERP de cada uno de los dos buscadores. Haga capturas de pantalla de cada SERP para mostrar al menos tres secciones destacadas en cada SERP. Añada algún comentario sobre la estructura o los componentes de la SERP de cada buscador. Indique cuál de las dos SERP (la de Google o la de Bing) podría resultar más útil a un periodista que necesita documentarse de forma rápida pero solvente sobre el tema utilizado. No tema ser subjetivo, aquí no hay forma de evitarlo, pero razone la respuesta.

1.2. Primera introducción al posicionamiento web

El cálculo de relevancia

Cada documento o cada página web presenta un grado de relevancia diferente para cada pregunta o necesidad de información. Ante una pregunta dada, como en “Premios Gaudí 2020”, en un extremo tendremos a las páginas cuya relevancia será cero (por tratar, por ejemplo, sobre Gaudí, pero no tener relación con el cine). En el otro extremo tendremos (con suerte) documentos cuya relevancia será total (o casi total) por tratar efectivamente de los Premios Gaudí del 2020 y con toda clase de información sobre nominados y premiados. En zonas intermedias tendremos documentos con diversos grados de relevancia.

Lo más lógico será presentar la página de resultados o SERP con los documentos ordenados según su grado de relevancia. Esta ordenación se denomina también *ranking*. De este modo, aunque haya miles de documentos relevantes obtenidos como respuesta, podremos limitarnos a consultar los 10 primeros, puesto que son los más relevantes. Otros modos de ordenación, como el alfabético o el cronológico no nos permitirían tal economía de consulta. Para estar seguros que accedemos a los más relevantes ¡tendríamos que examinar la lista completa aunque estuviera formada por millones de documentos!

¿Cómo saben los buscadores cómo deben ordenar los resultados para cada consulta? Nos interesa saber esto, porque sabiendo con qué criterios ordenan los resultados, sabremos no solamente buscar mejor, sino también como producir contenidos con más posibilidades de que tengan la máxima visibilidad (esto es, SEO periodístico).

Los buscadores actuales suelen combinar dos grupos de criterios para determinar la relevancia de una página web en relación a cada pregunta:

- Criterios *On page*
- Criterios *Off page*

Criterios internos – *On page*

Los criterios internos se refieren al contenido de la página. Entre estos criterios se encuentran la aparición de las palabras clave (de la pregunta) entre los contenidos de la página, así como la de sinónimos y términos relacionados. Es decir, a igualdad de otros aspectos, los buscadores consideran más relevantes (y pondrán en posiciones más destacadas) a las páginas que contengan más veces la palabra clave de la consulta y mayor número de sinónimos y términos relacionados de la misma. También interviene la extensión del documento (en general, más es mejor), así como el uso de componentes multimedia, la mención de fuentes (con enlaces), la correcta organización de la páginas, el uso de subtítulos, etc.

Otros aspectos, como que la palabra clave aparezca rodeada de etiquetas como <h1> o <h2> también otorgan mayor importancia relativa a la página correspondiente, y sobre todo, al hecho de que la palabra clave forme parte de la URL de la página. Algunos buscadores, como Google, indican que utilizan cientos de indicadores como los mencionados. Otros indicadores

incluyen la antigüedad del dominio, el volumen de información, la frecuencia de actualización, el uso de componentes multimedia, etc.

Criterios externos – Off Page

Los criterios externos se refieren, principalmente, al resultado que pueda arrojar el análisis de enlaces del sitio. El análisis de enlace incluye tanto los enlaces de salida (los enlaces de la página considerada hacia otras páginas) como los enlaces de entrada (los enlaces hacia la página considerada procedentes de otras páginas). De todas formas, el peso mayor recae en los enlaces de entrada. En este sentido, en general, una página es más importante cuantos más enlaces de entrada tiene procedentes de otras páginas.

Sin embargo, no todos los enlaces valen igual. Los enlaces de entrada de páginas que, a su vez, son muy enlazadas, son más importantes. Es decir, si la página **B** está enlazada por otras 10, y en cambio la página **A** está enlazada por otras 20, la página **A** es más importante que la **B**, siempre a igualdad de importancia de las páginas que envían el enlace. En cambio, si los enlaces que recibe la página **B** son de páginas, a su vez muy enlazadas, la página **B**, con menor número de enlaces puede ser considerada más importante que la **A**.

Cada buscador tiene sus propios conjuntos de criterios y sus propias reglas para asignar pesos a cada criterio, pero en general, el número de enlaces que recibe una página y la antigüedad de la misma (un valor positivo) suele ser los más importantes, al menos a igualdad de los otros factores.

SERP y snippets

Por último, la respuesta del motor con la lista de documentos organizados en un ranking debe presentarse de alguna forma. De este aspecto se ocupa precisamente la página de resultados o **SERP** (por *Search Engine Results Page*), que ya hemos tenido ocasión de revisar anteriormente con cierto detalle.

Como ya sabemos, una mirada atenta a la SERP de, por ejemplo, Google, nos muestra las siguientes informaciones:

1. Eventualmente, una lista de resultados de pago o **publicidad**.
2. La sección de resultados no pagados o resultados **orgánicos**.
3. De cada uno de estos resultados **orgánicos**, tenemos la siguiente información:
 - URL de la página
 - Título de la página (o del documento)
 - Fecha y dos líneas de descripción del contenido de la página
 - Eventualmente, otras informaciones, dependiendo del tipo de página y de su contenido.

Nos interesa mucho centrarnos ahora en el punto 3. El conjunto de datos que se presentan sobre cada página web se denomina **snippet** o fragmento. La siguiente captura muestra cuatro snippets tomados de una SERP sobre “desinformación”:

 es.greenpeace.org › en-profundidad › que-es-la-desinformacion ▼

¿Qué es la desinformación? - ES | Greenpeace España

21 may. 2019 - La **desinformación** no es nueva. A lo largo de los siglos los estamentos de poder han promulgado información más conforme a sus intereses ...

 www.politicaexterior.com › la-desinformacion-nueva-generacion ▼

La desinformación de nueva generación

17 sept. 2019 - Vivimos en realidades confusas. Los intentos de manipulación no tienen límites geográficos ni un único origen. Es la **desinformación** de nueva ...

 www.elheraldo.hn › opinion › columnas › el-poder-de-la-desinforma... ▼

El poder de la desinformación - Diario El Heraldo

Desinformar, según la Real Academia Española, es: Dar información intencionadamente manipulada al servicio de ciertos fines. También hace referencia a ...

 ec.europa.eu › Comisión Europea › España › Actualidad ▼

Desinformación: ¿cómo la detectamos? | España

21 may. 2019 - La **desinformación** nos llega todos los días por medios de comunicación y redes sociales, ¿sabes qué puedes hacer para detectarla y así ...

En la captura precedente vemos el ejemplo de cuatro snippets, uno por cada resultado. En la captura siguiente, analizamos la estructura concreta de uno de los snippets:

URL de la página 

 ec.europa.eu › Comisión Europea › España › Actualidad ▼

Título  **Desinformación: ¿cómo la detectamos? | España**

21 may. 2019 - La **desinformación** nos llega todos los días por medios de comunicación y redes sociales, ¿sabes qué puedes hacer para detectarla y así ... **Resumen** 

No es casualidad que la URL esté situada en la primera línea del snippet. Observar la URL es la primera medida contra el fraude y la desinformación cuando hacemos búsquedas en la web, porque la URL es difícil de trucar. En este caso, vemos que se trata de una información de la Unión Europea gracias a la URL que no deja lugar a dudas. Después, aparece el título, la fecha de publicación y un resumen de unas dos líneas.

El título y el resumen que aparecen aquí no están tomados necesariamente del título y del contenido de la página en cuestión, sino que proceden de los **metadatos** de la página, que en principio no son visibles para seres humanos, sino que se escriben en un apartado de la página que solo es visible para buscadores. Luego, los buscadores toman estos metadatos y

construyen el snippet. Lo veremos en su momento porque es clave para el SEO periodístico.

Snippets enriquecidos

Además de los snippets clásicos, a veces podemos ver snippets o resultados enriquecidos. Son fragmentos o snippets que añaden algún pequeño detalle o información extra a los tres componentes normales. La captura siguiente muestra un ejemplo de resultado enriquecido:

 www.filmaffinity.com › film895384 ▼

Dunkerque (2017) - Filmaffinity

★★★★★ Valoración: 7,1/10 - 46.204 votos

Dunkerque es una **película** dirigida por Christopher Nolan con Fionn Whitehead, Mark Rylance, Kenneth Branagh, Tom Hardy, Año: 2017. Título original: ...

[Críticas](#) · [Mostrar 49 críticas más](#) · [Fionn Whitehead](#) · [Tráilers](#)

En la captura anterior, podemos ver un ejemplo de rich snippet en el que aparecen valoraciones en la parte superior del resumen, y una serie de enlaces relacionados en la parte inferior.

Los rich snippets adoptan diversas formas según el tipo de resultado. Los más habituales son eventos (como estrenos cinematográficos), recetas, reseñas críticas, etc. Por ejemplo, en el siguiente caso vemos que el rich snippet de una receta incluye otros elementos, como foto, calorías y tiempo de preparación:

 www.recetasderechupete.com › tortilla-de-patatas-espanola-receta-pas... ▼

Tortilla de patatas y cebolla. Receta paso a paso

 ★★★★★ Valoración: 3,6 - 617 votos - 30 min - Calorías: 315

Cómo preparar una **tortilla de patatas** española. Receta fácil para que te salga como la de tu madre, perfecta. Preparación paso a paso, fotos y consejos.

Los resultados que Google muestra de este modo tienen más visitas. Por eso, los rich snippets son importantes para el SEO periodístico.

Actividad 5

Utilizando Bing, Qwant y Google, compare los diferentes snippets de cada buscador para la misma pregunta utilizando cualquier término de los que hemos usado hasta ahora. Haga capturas de pantalla con ejemplos de snippets de los tres buscadores y comente las posibles diferencias o semejanzas. Intente encontrar diferentes ejemplos de snippets si es posible (algunas búsquedas no producen rich snippets, si es el caso, intente hacer búsquedas por títulos de filmes o de eventos culturales. Si aún así no consigue resultados con rich snippets, puede mostrar en las capturas los resultados normales).

Conclusiones

Conocer los aspectos estructurales y funcionales que hemos examinado aquí nos ayudará sin ninguna duda en nuestra relación con los buscadores. Conocer un poco mejor los fundamentos de los motores nos ayudará a llevar a cabo mejores búsquedas como periodistas porque entendemos un poco mejor cómo "piensan" los buscadores y porqué determinados resultados están situados más arriba en el ranking que otros. Adicionalmente, si en nuestras responsabilidades entra publicar contenidos en la web entenderemos mejor qué debemos hacer para nuestras páginas aparezcan con la máxima visibilidad posible en los buscadores.

2. Búsqueda avanzada

Introducción

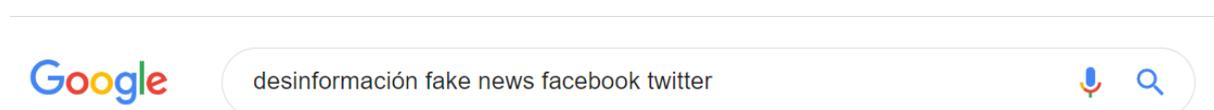
Afortunadamente, hasta un niño sabe utilizar la búsqueda simple. Los buscadores se han esforzado, siguiendo la estela de Google en ofrecer una opción de búsqueda simple realmente *simple*. Escribimos una palabra (o una frase) en la caja de búsqueda y después de “hacer intro”, ya está. Es realmente importante que la búsqueda simple sea efectivamente muy fácil de llevar a cabo si no, no merecería ese nombre.

Sin embargo, para búsquedas especializadas o para búsquedas en un contexto profesional como el de los profesionales de la comunicación, o en un contexto académico (en vez de ocio, por ejemplo) es necesario utilizar la búsqueda avanzada para obtener resultados de calidad y poder completar nuestra misión, sea la que sea, si incluye obtener información.

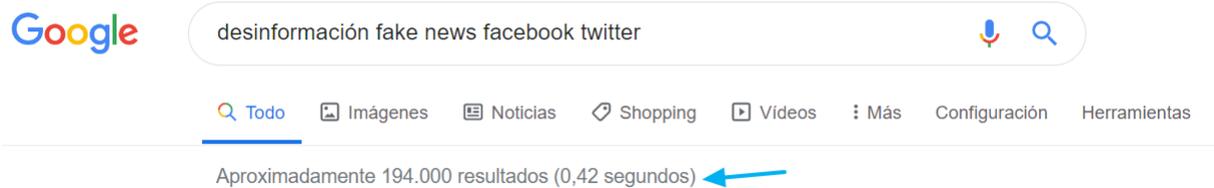
Ecuaciones de búsqueda

“Ecuaciones de búsqueda” es un término técnico heredado de las primeras bases de datos. Actualmente es trivial entrar uno o más términos en un buscador, lanzar una consulta y obtener con facilidad algunos resultados relevantes, de modo que referirse a esa operación como “ecuación de búsqueda” suena innecesariamente técnico (o peor, pedante), sin embargo, tiene sentido mantener el nombre para entender lo que sigue.

No es que las **ecuaciones de búsqueda** hayan dejado de existir o de tener importancia, es que su papel nos pasa simplemente desapercibido. Técnicamente, cuando entramos una palabra clave o una frase clave en la caja de búsqueda de Google o de Bing estamos usando una ecuación de búsqueda, aunque sea muy simple:



En la captura precedente vemos una ecuación de búsqueda formada por cinco palabras: [desinformación fake news Facebook twitter]. Decimos que es una ecuación porque tiene una solución, que consiste en la SERP, o lista de resultados que cumplen la ecuación. Según la captura ampliada que mostramos a continuación, cumplen esta ecuación casi 200 mil documentos:



Ahora bien, si reflexionamos lo que este conjunto de cinco palabras pueden significar, seguramente encontramos que, la necesidad de información de nuestro usuarios debe ser algo como esto (expresado en lenguaje natural):

(1)

“Necesito datos sobre desinformación, incluyendo el tema de las llamadas fake news, y su relación con Facebook y Twitter”

Teniendo lo anterior en cuenta, una ecuación de búsqueda mejor formada sería la siguiente:

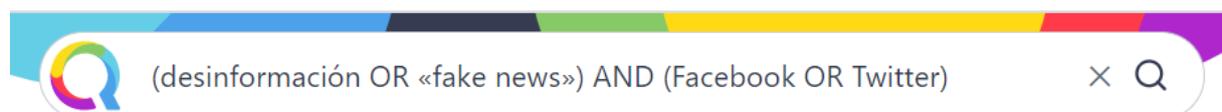
(2)

(desinformación OR “fake news”) AND (Facebook OR Twitter)

Ahora bien, para pasar de (1) a (2) necesitamos conocer cómo se forman las ecuaciones de búsqueda, en cuyo caso, podremos ponerlas de este modo en un buscador determinado como Google:



O de este modo, en otro buscador determinado como Qwant:



Lo que conseguiremos es expresar de forma mucho más adecuada nuestra necesidad de información y de este modo conseguiremos que los primeros resultados resulten mucho más precisos.

Aún no sabemos qué papel juegan los llamados *operadores booleanos* como OR y AND, ni porqué hemos usado comillas en “fake news”, pero todo esto lo vamos a examinar a continuación mediante un conjunto de tablas.

Tabla 1: Tipos de ecuaciones de búsqueda más usuales

<i>Tipo</i>	<i>Ejemplos</i>	<i>Explicación</i>
Simple	cine	Busca documentos que contengan la cadena de caracteres indicada, en este caso, “cine”.

Frase	"john ford"	Fuerza al sistema para que la cadena de caracteres buscada sea exactamente como aparece entre las comillas. En este caso, primero la palabra john, después un espacio en blanco y después la palabra ford sin nada más entre ellas. Evita las falsas coordinaciones. Por ejemplo, la página de una empresa llamada John Smith que publica información de los coches de la marca Ford.
Booleana simple	televisión AND noticiarios cine OR televisión cine NOT western	AND: exige que ambas palabras estén presentes en el documento. En algunos buscadores, como Google, un espacio en blanco equivale a un AND. En el ejemplo, buscamos documentos que hablen de televisión y también de noticiarios OR: al menos una de las palabras debe estar presente; pueden estar presentes las dos. En el ejemplo, buscamos documentos que hablen o bien de cine o bien de televisión o bien de ambos. NOT: documentos donde la primera palabra debe estar presente, siempre que no aparezca la segunda palabra. En el ejemplo, todos los documentos que tengan la palabra cine, siempre que no tengan la palabra western.
Booleana compleja	(cine OR televisión) AND (política OR economía)	Combina varias palabras clave con dos o más operadores booleanos. Los paréntesis se pueden utilizar para delimitar de forma exacta el orden y el alcance de cada operador. En este caso: primero se crea el conjunto 1 con todos los documentos que tienen la palabra cine, la palabra televisión o ambas. Después se crea el conjunto 2 con todos los documentos que tienen la palabra política, la palabra economía o ambas. Por último, se combinan los resultados anteriores con un AND y se produce como respuesta el conjunto de documentos que tratan o bien de cine o bien de televisión, siempre que además traten o bien de política o bien de economía

Además de las búsquedas anteriores, algunos buscadores (como Google especialmente) ofrecen una gran variedad de posibilidades que a veces tiene poco que ver en realidad con la recuperación de información. Por ejemplo, el lenguaje de interrogación de Google actúa también como una calculadora aritmética si entramos expresiones de suma, resta, etc. (p.e. $9*670/8$). Otras características se asemejan a lo que podríamos llamar un sistema de respuestas directas, es decir, no un sistema que busca documentos que contengan la respuesta, sino un sistema que responde a la pregunta.

Por ejemplo, imagine que entramos la siguiente expresión en Google:

- moneda de Jordania

y como respuesta tenemos:

- Dinar jordano

en lugar de una lista de (solamente) documentos que tienen la expresión "moneda de

Jordania”.

Actividad 6

Entre en esta página de Google: <https://support.google.com/websearch/answer/2466433>
Revise la lista de tipos de búsqueda que Google explica que se pueden hacer. Ponga a prueba al menos dos de las que le parezcan más interesantes o de las que entienda mejor su significado. Si de alguna clase de búsqueda no entiende su función exacta, no se preocupe, puede ignorarla de momento. Haga capturas de pantalla de sus pruebas.

Tabla 2: Búsqueda parametrizada

<i>Variación</i>	<i>Ejemplo</i>	<i>Explicación</i>
Genérica	cine	Busca la cadena de caracteres en cualquier lugar del documento, en documentos de cualquier fecha, en todos los formatos de documento, etc.
Parametrizada	allintitle:cine filetype:pdf desinformación site:edu	<p>Busca la cadena de caracteres en el campo título (allintitle:cine) y solamente en documentos con formato pdf (filetype:pdf). Otro ejemplo muy importante es poder buscar en determinados sitios, por ejemplo, sitios académicos con el operador site, mediante una búsqueda tipo [desinformación site:edu]. Estos nos permitiría encontrar informes académicos sobre desinformación.</p> <p>La sintaxis del ejemplo es la de Google. Otros buscadores pueden utilizar otras expresiones. Acotaciones principales:</p> <ul style="list-style-type: none">Por campos (p.e. título)Por fechas (p.e. Últimos 6 meses)Por tipo de documento (p.e. pdf)Por idioma (p.e. español)Por dominio (p.e: gov o edu)

Tabla 3: Interrogación no booleana vs. booleana

<i>Tipo de interrogación</i>		<i>Significado lógico</i>
No booleana (sin operadores explícitos)	cine literatura niños adolescentes	Documentos que contengan todas y cada una de las palabras. No capta la necesidad real del usuario. Se pierden de facto, por ejemplo, documentos que traten <i>solamente</i> de cine y de adolescentes. También los que traten <i>solamente</i> de cine y de niños, etc; puesto que pedimos exactamente que todas, es decir, las cuatro palabras estén presentes.
Booleana (con operadores explícitos)	(cine OR literatura) AND (niños OR adolescentes)	Capta exactamente la necesidad de información del usuario. Por ejemplo, un documento sobre niños y literatura será recuperado, aunque no hable ni de adolescentes ni de cine, etc., y esto es exactamente lo que necesita el usuario.

Como se puede ver, solamente la forma booleana capta de forma adecuada la semántica de esta necesidad de información, esto es, solamente la fórmula booleana es capaz de evitar tanto el ruido (información no deseada) como el silencio (información no recuperada).

El principal inconveniente es que el álgebra booleana es, de facto, anti intuitiva e intimidatoria para el usuario final. Por ejemplo, si revisamos la representación de la necesidad de información anterior:

cine **y** literatura **para** niños **y** adolescentes

la mayor parte de los usuarios que se hubieran visto obligados a traducirla a una expresión booleana lo hubieran hecho así:

cine **AND** literatura **AND** niños **AND** adolescentes

Por el simple procedimiento de traducir todos los **y** gramaticales en **AND** lógicos (incluida la expresión **para**), y que es obviamente incorrecta por las razones señaladas antes, pero que vamos recordar de nuevo: **no** estamos interesados **solo** en documentos que traten **todos** estos aspectos **en el mismo documento**; sino, literalmente en documentos que traten o bien de niños y cine, o bien de niños y literatura, o bien de adolescentes y cine, o bien de adolescentes y literatura, y ya sea por separado o conjuntamente.

Formularios de búsqueda

Por suerte, algunos buscadores, como Google, ayudan a los usuarios a formular lo mejor posible sus ecuaciones de búsqueda, incluso sin conocer las bases teóricas (aunque conocerlas ayuda mucho), gracias al uso de formularios. En concreto, Google utiliza uno que ya vimos anteriormente, pero por facilidad volvemos a reproducir:

Búsqueda avanzada

Buscar páginas con...

todas estas palabras:

esta palabra o frase exactas:

cualquiera de estas palabras:

ninguna de estas palabras:

números desde el:

hasta

Haz lo siguiente en el cuadro de búsqueda

Escribe las palabras importantes: terrier ratonero tricolor

Escribe las palabras exactas entre comillas: "terrier ratonero"

Escribe OR entre todas las palabras que quieras: miniatura OR estándar

Añade un signo menos delante de las palabras que no quieras que aparezcan: -roedor, -"Jack Russell"

Escribe dos puntos seguidos entre los números y añade una unidad de medida: 10..35 kg, 300..500 euros, 2010..2011

A continuación, limitar los resultados por...

idioma:

Busca páginas en el idioma que selecciones.

región:

Busca páginas publicadas en una región determinada.

última actualización:

Busca páginas actualizadas en el periodo de tiempo especificado.

sitio o dominio:

Busca un sitio (como wikipedia.org) o limita los resultados a un dominio como, por ejemplo, .edu, .org o .gov.

los términos que aparecen:

Busca términos en toda la página, en el título de la página o en la dirección web, o enlaces a la página que estás buscando.

Búsqueda Segura:

Indica a **Búsqueda Segura** si quieres que filtre contenido sexualmente explícito.

tipo de archivo:

Busca páginas en el formato que prefieras.

derechos de uso:

Busca páginas que puedas utilizar libremente.

[Búsqueda avanzada](#)

También puedes...

[Buscar páginas similares a una URL](#)[Buscar páginas visitadas](#)[Usar operadores en el cuadro de búsqueda](#)[Personalizar configuración de búsqueda](#)

España

En la captura vemos que el formulario de búsqueda avanzada de Google, que puede utilizarse aquí: https://www.google.com/advanced_search, tiene dos claras secciones:

- El apartado para entrar palabras clave > para las búsquedas booleanas.
- El apartado para establecer filtros > para las búsquedas parametrizadas.

El siguiente ejemplo muestra un ejemplo de utilización del formulario:

Búsqueda avanzada

Buscar páginas con...

todas estas palabras:

desinformación periodismo

esta palabra o frase exactas:

cualquiera de estas palabras:

ninguna de estas palabras:

números desde el:

hasta

A continuación, limitar los resultados por...

idioma:

cualquier idioma

región:

cualquier región

última actualización:

en cualquier momento

sitio o dominio:

los términos que aparecen:

en el título de la página

Búsqueda Segura:

Ver resultados más relevantes

tipo de archivo:

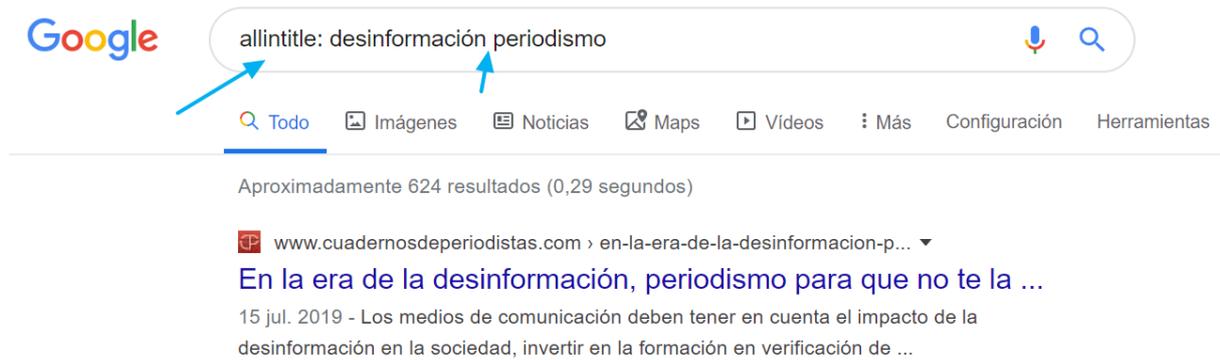
cualquier formato

derechos de uso:

sin filtrar por licencia

Búsqueda avanzada

Una vez lanzada la anterior búsqueda, produce esta ecuación en la página de resultados de Google:



En la captura precedente, vemos que Google ha añadido el operador “allintitle” delante de las palabras clave, así como el espacio en blanco entre las dos palabras clave significa que están combinadas con un AND. Los resultados así son mucho más precisos.

Actividad opcional: mediante pruebas libres, intente determinar la forma de llevar a cabo búsquedas mediante el formulario de búsqueda avanzada de Google. Pruebe a combinar algunas palabras clave con algunos filtros. Puede hacer pruebas de modo libre. Si lleva a cabo esta práctica, puede documentarla con algunas capturas de pantalla a su criterio. Muy importante: la ayuda que aparece al lado de cada opción es para cuando estamos usando la caja de búsqueda simple, no para este formulario.

Conclusiones

En el contexto de las bases de datos de prensa, así como en bancos de imágenes y video, el uso de búsquedas booleanas es imprescindible. En cambio, en buscadores, su utilidad es menos evidente porque casi cualquier búsqueda, incluso las peor formuladas arroja algún resultado.

Es decir, al usuario típico de un buscador que busca información por motivos de ocio o por simple curiosidad no le importa perder documentos relevantes mientras que los 10 o 20 primeros, que son los que estará dispuesto a ver como respuesta a su búsqueda, sean relevantes y solucionen (mejor o peor) su problema de información.

En cambio, para búsquedas profesionales el conocimiento de la búsqueda avanzada es imprescindible para resolver con éxito misiones críticas. En especial, en el campo del periodismo y de la comunicación audiovisual es imprescindible saber utilizar operadores como los que presenta la búsqueda avanzada. Poder limitar resultados por tipo de documento, por servidor donde ha sido publicado, limitar resultados por el título, etc., puede marcar la diferencia entre el éxito o el fracaso en una misión que implique buscar o investigar en internet.

3. Evaluación de sitios web

Finalmente, tanto los periodistas como el profesional de la comunicación están obligados a verificar sus fuentes, y esto es así tanto si por “fuentes” entendemos personas como sitios web (o cualquier clase de recursos digitales), como es el caso de esta Unidad.

3.1. Dimensiones de la calidad

Desde el punto de vista de la documentación periodística, la calidad de un sitio web presenta al menos dos parámetros básicos

1. Contenido
2. Credibilidad

3.1.1. Contenido

Los indicadores que nos ayudan a determinar la calidad de los contenidos de un sitio son, entre otros los siguientes:

1. Rigor. Se refiere al cuidado con el que ha sido preparada la información. No siempre podemos estar seguros respecto a esto, y con frecuencia debemos guiarnos por indicios más o menos (in)directos. Algunos de estos indicios son: la mención de otras fuentes, la mención de datos concretos (fechas, autores, estadísticas), la ausencia de errores factuales, la ausencia de juicios de valor gratuitos, la separación cuidadosa entre opinión e información, etc., por no mencionar la ausencia de errores ortográficos, etc.

Un buen ejemplo de esto suelen ser sitios web de gran relieve en el mundo de las políticas económicas, como el sitio web del Banco Mundial (www.bancomundial.org) o el sitio de la OCDE (<http://www.oecd.org>) o en sitios que ejercen funciones de “watchdog” como el de Human Right Watch.

2. Exhaustividad. Este indicador se refiere al grado en el cual las informaciones se presentan de modo más o menos completo. Por ejemplo, un dato bibliográfico puede consistir solamente en un título y un nombre de autor, o incorporar también la fecha de edición, el lugar de edición, el nombre de la editorial, el número de páginas de la obra y el número de ISBN. Un directorio de empresas puede dar un nombre y una ciudad para cada empresa o la ficha completa de la misma, incluyendo el nombre de los miembros del consejo de administración; una biografía puede dedicar un párrafo a cada personaje o el equivalente a varias páginas de información, etc.

Un excelente ejemplo de exhaustividad en la mayor parte de sus informaciones la encontramos en la base de datos de cine IMDB

3. Actualización. Señala la frecuencia o la fecha de la última actualización. Es evidente que, en general y a igualdad de otros factores, cuanto más actualizada sea una información, mayor valor tiene. Si consultamos un atlas geográfico de centro Europa desearemos, con toda seguridad, que sea lo más actualizado posible.

En general, los sitios web de medios diarios, son ahora buenos ejemplos de una política adecuada de actualización, con varias ediciones diarias.

4. Edición. Este indicador se refiere al grado, tanto explícito como implícito, y más o menos evidente según los casos, en el que la información ha sido tratada, revisada, completada, editada en una palabra, antes de ser publicada. Ofrece evidencias de edición una web que presenta un título en cada página, un nombre de autor (o una indicación de autoría) y una fecha de creación, por no mencionar, una vez más, la ausencia de errores ortográficos o la simple presentación descuidada. Las menciones de autor, la titulación y la datación deben repetirse en todas las secciones de la web si procede, es decir, si la naturaleza del recurso lo exige. Por ejemplo, la web de una publicación periódica deberá indicar no solamente los datos del organismo editor, el nombre del director o directora de la publicación, etc., sino que deberá presentar esos datos para cada artículo de la publicación, etc.

Desde hace algún tiempo, los proyectos de la organización Wikimedia, entre los que se incluye la Wikipedia son un buen ejemplo de edición cuidada, donde varios editores supervisan que los artículos de la wikipedia dispongan de datos verificados. Incluso cuando no es así, se señala la carencia de ellos para alertar al lector.

5. Sistematización. El tratamiento sistemático proporciona una gran calidad a la información, por eso debe considerarse un indicador de primera magnitud. En el punto anterior hemos mencionado un factor de sistematización, pero cada tipo de recurso tiene los suyos propios. Por ejemplo, en un directorio hay sistematización si cada web está descrita de la misma forma. Hay sistematización en una base de datos cuyos registros poseen toda la misma estructura básica, etc.

Las bases de datos que hemos revisado en esta Unidad cumplen todas con esta característica. Pero los principales diarios en línea también, ya que procuran presentar las noticias siempre con un tratamiento uniforme.

6. Interés intrínseco. Aunque sea una obviedad, hay que recordar que existen temas completamente triviales, y muchas webs dedicadas a ellos. De hecho, la trivialidad y el esnobismo parecen ejercer una atracción fatal en el mundo de la WWW. Dicho de otro modo, debe considerarse el interés intrínseco de la información por encima de consideraciones de estética o de esnobismo. En este sentido, otro motivo de sorpresa es la enorme atención que algunas webs de contenidos voluntariamente triviales merecen en los medios de comunicación.

Una vez más, la Wikipedia o cualquier sitio de los que hemos comentado en esta Unidad cumplen este criterio.

7. Originalidad. Este parámetro se refiere al grado en el cual un recurso ofrece informaciones únicas o exclusivas en algún sentido. Muchas webs ofrecen informaciones relacionadas con temas de actualidad, pero solamente algunas webs ofrecerán, además, tratamientos en profundidad de tales temas o recursos complementarios, como gráficos, mapas, infografías, datos estadísticos, puntos de vista opuestos, etc. Muchas webs ofrecen listas de recursos digitales, pero solamente algunas evalúan esos recursos, etc. Este indicador mide, en definitiva, el grado en el cual un recurso ofrece algo que únicamente puede encontrarse en el mismo.

3.1.2. Credibilidad

Los indicadores anteriores son directamente observables o medibles y, en todo caso, relativamente fáciles de determinar.

El cambio, la credibilidad se refiere a una valoración donde, además de un inevitable componente subjetivo por parte del analista, tenemos que contar con que los responsables del sitio suelen estar interesados en parecer creíbles, independientemente de que lo sean o no.

Determinar la credibilidad de un sitio puede requerir indagaciones indirectas, así como efectuar inferencias sin garantía de acierto al cien por cien. Por último, la indagación puede no aportar ningún dato concluyente. Dicho de otro modo, mientras que siempre podremos saber o decir algo sobre el contenido, puede que no siempre podemos llegar a conclusiones seguras sobre la credibilidad.

No obstante, es éste último (la credibilidad) el aspecto que más debe preocupar en el contexto del trabajo periodístico por razones evidentes e independientes de su facilidad o dificultad de determinación.

En todo caso algunos indicadores relativamente fáciles de determinar son los siguientes:

Por su parte, el parámetro de la autoría puede evaluarse a través de tres aspectos o indicadores diferentes (aunque muy relacionados entre sí, claro está):

1. *Página de créditos.* La existencia misma de declaraciones explícitas de autoría es ya una primera indicación de credibilidad. Sorprende el elevado número de webs que carecen de indicaciones claras sobre a quién cabe atribuir la paternidad, la propiedad o la responsabilidad intelectual de la misma. Por eso, de entrada, es ya un indicador de calidad la simple presencia de esa información bajo apartados denominados *Créditos*, *Quiénes somos*, *About Us*, etc. Una web bien concebida debe tener un apartado donde se explique sin ambigüedad y con el máximo detalle posible: a quién o a quienes debe atribuirse la responsabilidad del recurso; quién o quiénes son los editores o distribuidores del recurso. Esta declaración debería ir acompañada de datos complementarios sobre los autores y los editores: institución donde trabajan; dirección postal de los editores, etc. Por el contrario, una web sin tales indicaciones debe considerarse, de entrada, poco solvente.

Un sitio que puede ser un buen ejemplo de cómo resuelve este apartado de forma ejemplar es la página de créditos de la televisión pública de Estados Unidos: PBS (www.pbs.org).

3. *Solvencia o adecuación de la fuente.* La solvencia del autor y/o de la institución que ha creado, producido o editado el recurso. Es evidente que no merece el

mismo crédito una web sobre salud producida por un doctor o una doctora en Medicina que por un curandero, para poner un ejemplo de los llamados "de manual". Por tanto, un indicador de calidad en este sentido será la adecuación curricular del autor al tema tratado. Por parecido razonamiento, es evidente que no podemos otorgar el mismo crédito a un informe sobre política cultural publicado en la web de la Unesco que uno publicado en la web de un grupo que haga llamamientos al exterminio de media humanidad. Por tanto, respecto a la figura del editor o distribuidor, su solvencia pública será otro indicador de importancia capital. Como regla general, es evidente que existen organismos que ofrecen mayores garantías, de entrada, que otros. Por ejemplo, siempre suelen ser un buen indicador de calidad el que un recurso haya sido publicado por una universidad, o por una editorial de prestigio, etc.

Por ejemplo, la web del Banco Mundial o de la OCDE, ya indicadas anteriormente son buenos ejemplos de fuentes solventes por lo que hace a datos e informaciones económicas.

Ahora bien, ¿qué sucede cuando no existe en el sitio una indicación clara de autoría? O, peor aún, ¿qué sucede cuando tenemos sospechas sobre la autoría o la identidad declarada del sitio?

Ante la situación anterior, tenemos una pequeña batería de recursos que nos proporciona la propia web para intentar determinar "quién está detrás" de un sitio determinado.

Antes de presentar estos recursos, no obstante, una observación. Lo más operativo cuando nos encontramos ante un sitio del que no estamos seguros de su fiabilidad o credibilidad y su página de créditos no nos ayuda a despejar dudas consiste simplemente en ignorarlo y buscar otra fuente. En la web, por suerte, lo que *abunda* especialmente es la información, así que en general siempre tendremos alternativas donde escoger si una fuente no merece nuestra confianza.

Por tanto, lo que sigue solamente tiene sentido si, por alguna razón, y a pesar de todo estamos especialmente interesados en saber más sobre un sitio. Las herramientas que podemos utilizar son las siguientes (ninguna de ellas, sin embargo es concluyente):

1. **Alexa.** Este directorio (www.alexacom.com) proporciona información muy importante sobre cualquier sitio. Nos indica: (i) el tráfico del mismo (el número de visitas que recibe); (ii) otros sitios que enlazan al sitio analizado; (iii) comentarios o revisiones sobre el sitio (si existen) y otras indicaciones estadísticas muy importantes sobre el sitio, permitiendo incluso hacer comparaciones entre varios sitios. Es muy importante el perfil que Alexa nos proporciona así de un sitio. Si apenas recibe visitas (o al contrario), o si quienes lo enlazan tienen un sesgo determinado, etc., en conjunto podremos hacernos

una idea bastante determinada de la credibilidad del sitio. Naturalmente, no es un sistema infalible. En general, sirve más para confirmar la credibilidad de un sitio que para lo contrario, porque sitios poco visitados y poco enlazados no aparecen en su directorio. No obstante, la ausencia de un sitio en Alexa también puede ser (no necesariamente) un elemento de advertencia. Siempre puede ayudar saber qué sitios enlazan al sitio considerado. Por ejemplo, un sitio sobre salud que recibe enlaces principalmente de sitios vinculados con la parasicología o procedentes de una secta, etc., naturalmente no es un buen indicio (tampoco es infalible, un sitio puede ser enlazado sin permiso de los responsables del mismo).

2. **Servicios Whois.** Existe una serie de sitios que permiten lanzar consultas a los registros oficiales de los propietarios de los dominios. Un ejemplo es Whois.net (<http://www.whois.net/>). No son datos concluyentes, pero típicamente podemos saber el nombre de la empresa o de la persona que compró el dominio y sus datos básicos de contacto. No obstante, a veces están ocultos.
3. **Open Site Explorer** (<https://moz.com/researchtools/ose/>). Es un buscador especializado que analiza la URL que entremos y nos indica informaciones básicas sobre el sitio, algunas son en realidad muy similares a las que proporciona Alexa (qué otros sitios enlazan al sitio determinado, p.e.) pero siempre nos servirá para ampliar o comparar resultados y en teoría nos proporciona información de cualquier sitio aunque no forme parte del top 100 mil como en el caso de Alexa. **Alternativa: Ahref** (<https://ahrefs.com/>).
4. **Búsquedas en la web.** Naturalmente, siempre tenemos la opción de entrar el nombre y/o la URL del sitio en la caja de búsqueda de Google y ver qué dice el resto de la web (si es que dice algo) sobre el sitio en cuestión. Si no dice nada, ya es de por sí motivo de sospecha.
5. **Uso de navegadores de última generación.** Aunque parezca una trivialidad un periodista siempre debería utilizar navegadores de última generación, o al menos de versiones muy recientes. La razón es simple: cada nueva generación de los principales navegadores (IE, Firefox, Opera) añade mejores capacidades para detectar sitios fraudulentos o que usurpan la identidad de sitios reales, etc.

3.4. Conclusiones

Determinar la calidad de un sitio web puede ser absolutamente trivial. No hace mucho análisis para observar la enorme calidad de sitios como los del *New York Times*, *The Guardian*, *PBS*, *El País*, etc. Otras veces, tropezamos con páginas infames que inmediatamente nos advierten sobre su falta total de credibilidad.

Sin embargo, el problema nunca son este tipo de sitios que se sitúan en los extremos, sino aquellos que encontramos a través de una búsqueda en la web, de los que antes no sabíamos nada y presentan una apariencia como mínimo correcta. En estos sitios podemos encontrar información aparentemente útil y sin embargo tener dudas fundadas sobre su solvencia.

Si observamos unas pautas elementales, en realidad de sentido común, podemos evitar ser embaucados por sitios que no merecen nuestra confianza. En este apartado hemos proporcionado además algunas pautas que podemos seguir si, por algún motivo, es importante para nosotros saber más sobre un sitio y asegurarnos de su credibilidad o solvencia sobre los temas que trata.

Actividad 7

Haga una búsqueda en Google (o en Bing o Qwant) sobre un tema interés social, político o cultural de gran relevancia. Seleccione los tres primeros sitios de la primera página de resultados de Google y compare las respectivas páginas de créditos de cada uno de los sitios. Haga capturas de pantalla comentadas.

4. Referencias

- Abadal, Ernest; Codina, Lluís. 2005. *Bases de datos documentales: características, funciones y método*. Madrid: Síntesis, 220 p.
- Charlton, G. (2016) "How much do journalists and editors need to know about SEO?". Search Engine Watch. <https://searchenginewatch.com/2016/03/15/how-much-do-journalists-and-editors-need-to-know-about-seo/>
- Codina, L.; Iglesias-García, M.; Pedraza R.; García- Carretero, L. (2016). "Visibilidad y posicionamiento web de informaciones periodísticas: el framework SEO-RCP". Barcelona: Serie editorial DigiDoc-UPF. <https://repositori.upf.edu/handle/10230/26040>
- Codina, L.; Gonzalo-Penela, C.; Pedraza-Jiménez, R. y Rovira, C. (2017). Posicionamiento y medios de comunicación. Ciclo de vida de una campaña y factores SEO. Barcelona: Departamento de Comunicación. Serie Editorial DigiDoc. https://repositori.upf.edu/bitstream/handle/10230/28134/codina_2017_posi.pdf?sequence=1&isAllowed=y
- Codina, L.; Iglesias, M.; Pedraza, R.; García-Carretero, L. (2016). Visibilidad y Posicionamiento Web de informaciones periodísticas: El Framework SEO-RCP. Barcelona: Serie Editorial DigiDoc-UPF,

2016. https://repositori.upf.edu/bitstream/handle/10230/26040/codina_RCP_042016.pdf
- García-Carretero, L.; Codina, L.; Díaz-Noci, J. y Iglesias-García, M. (2016). Herramientas e indicadores SEO: características y aplicación al análisis de cybermedios. *El Profesional de la Información*, 25(3), pp. 497-504. <http://www.elprofesionaldelainformacion.com/contenidos/2016/may/19.pdf>
 - García-Carretero, Lucía y Codina, Lluís (2018). Visibilidad web y Comunicación Política: análisis de los sitios web de la coalición electoral Barcelona en Comú. En VI Congreso Internacional de la Asociación Española de Investigación en Comunicación AE-IC «Comunicación y Conocimiento» (pp. 2318-2347). Salamanca. ISBN: 978-84-09-03393-5. <https://ae-ic.org/final/Libro%20de%20Comunicaciones%20del%20VI%20congreso%20AE-IC.pdf>
 - Giomelakis, D.; Veglis, A. (2015). Employing Search Engine Optimization techniques in Online News. *Studies in Media and Communication* Vol. 3, No. 1; June 2015. <http://dx.doi.org/10.11114/smc.v3i1.683>
 - Giomelakis D.; Veglis A. (215a) "Investigating Search Engine Optimization Factors in Media Websites, The case of Greece". *Digital Journalism*, pp. 379-400 <https://doi.org/10.1080/21670811.2015.1046992>
 - Gonzalo, C.; Codina, L. y Rovira, C. (2015). Recuperación de información centrada en el usuario y SEO: categorización y determinación de las intenciones de búsqueda en la web. *Index comunicación*, 5 (3), pp. 19-27. <https://dialnet.unirioja.es/descarga/articulo/5405092.pdf>
 - Linares, J.; Codina, L.; Váñez, M. y Rodríguez-Martínez, R. (2016). Interactividad, buscabilidad y visibilidad en cybermedios: sistema de análisis y resultados. Barcelona: Universitat Pompeu Fabra. Departament de Comunicació https://repositori.upf.edu/bitstream/handle/10230/26280/Linares_inter_2016.pdf?sequence=1&isAllowed=y
 - Lopezosa, C. y Codina, L. (2018). Análisis del posicionamiento en medios de comunicación con herramientas SEO: cobertura informativa de los premios Oscar 2017. Barcelona: Universitat Pompeu Fabra. Departamento de Comunicación (Serie Editorial DigiDoc; EPI nº 2). <https://repositori.upf.edu/handle/10230/33632>
 - Lopezosa, C.; Codina, L. y Caldera-Serrano, J. (2018). "SEO semántico: Framework ISS para la optimización de sitios intensivos en contenidos", *Cuadernos de Documentación Multimedia*, 29, 97-123. <https://revistas.ucm.es/index.php/CDMU/article/view/60607/4564456547455>
 - Lopezosa, C.; Codina, L., Freixa, P. (2018). SEO y comunicación audiovisual: análisis comparativo de portales de vídeo bajo demanda. Barcelona: Digidoc-EPI. http://www.elprofesionaldelainformacion.com/documentos/seo_comunicacion_audiovisual_analisis.pdf
 - Lopezosa, C.; Codina, L.; Rovira C. (2019). Visibilidad web de portales de televisión y radio en España: ¿qué medios llevan a cabo un mejor posicionamiento en buscadores?. Barcelona: DigiDoc-EPI. https://repositori.upf.edu/bitstream/handle/10230/36234/Codina_epi_visi.pdf?sequence=1&isAllowed=y

- Lopezosa, C.; Orduña-Malea, E. y Pérez-Montoro, M. (2019). “Making Video News Visible: Identifying the Optimization Strategies of the Cybermedia on YouTube Using Web Metrics”, Journalism Practice, <https://www.tandfonline.com/doi/full/10.1080/17512786.2019.1628657>
- Pérez-Montoro, M.; Codina, L. (2017). Navigation Design and SEO for Content-Intensive Websites: A Guide for an Efficient Digital Communication. Oxford: Chandos Publishing (Elsevier). <https://www.sciencedirect.com/book/9780081006764/navigation-design-and-seo-for-content-intensive-websites>
- Smyrnaio, N; Sire, G. (2014) “The News according to Google How Does Algorithmic Infomediatio Frame the Work of French Journalists?”. Presented at JSS-ECREA 2014 Conference, Thessaloniki: <https://es.slideshare.net/smyrnaio/the-news-according-to-google-how-does-algorithmic-infomediatio-frame-the-work-of-french-journalists>
- Wordtracker (2018). “A guide to SEO for editorial teams and journalists”. Wordtracker. <https://www.wordtracker.com/academy/seo/getting-started/seo-for-editorial-teams-journalists>

Forma de citación de este documento: Lluís Codina, Rafael Pedraza, Cristòfol Rovira, Carlos Lopezosa. *Búsqueda y evaluación de la información para periodistas y comunicadores*. Barcelona: Grupo DigiDoc. Departamento de Comunicación. Universitat Pompeu Fabra, 2020.

Esta obra está publicada bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)
