

2. Los motores de búsqueda: funcionamiento

Este bloque explica la estructura y funcionamiento de los motores de búsqueda para internet.

- [2.1. Captura de información](#)
- [2.2. Procesamiento y recuperación de información](#)
- [2.3. Presentación de resultados](#)

2.1. Captura de información

La identificación y captura de información es llevada a cabo por el **robot** o *spider*. **El robot de búsqueda es una herramienta de software que rastrea internet, con la finalidad de localizar e identificar nuevas páginas y documentos**, y la de detectar si se han producido cambios o modificaciones en otras que ya hubiese localizado previamente. Las nuevas páginas encontradas son utilizadas, a su vez, como punto de partida para descubrir nueva información, utilizando los enlaces que contienen.

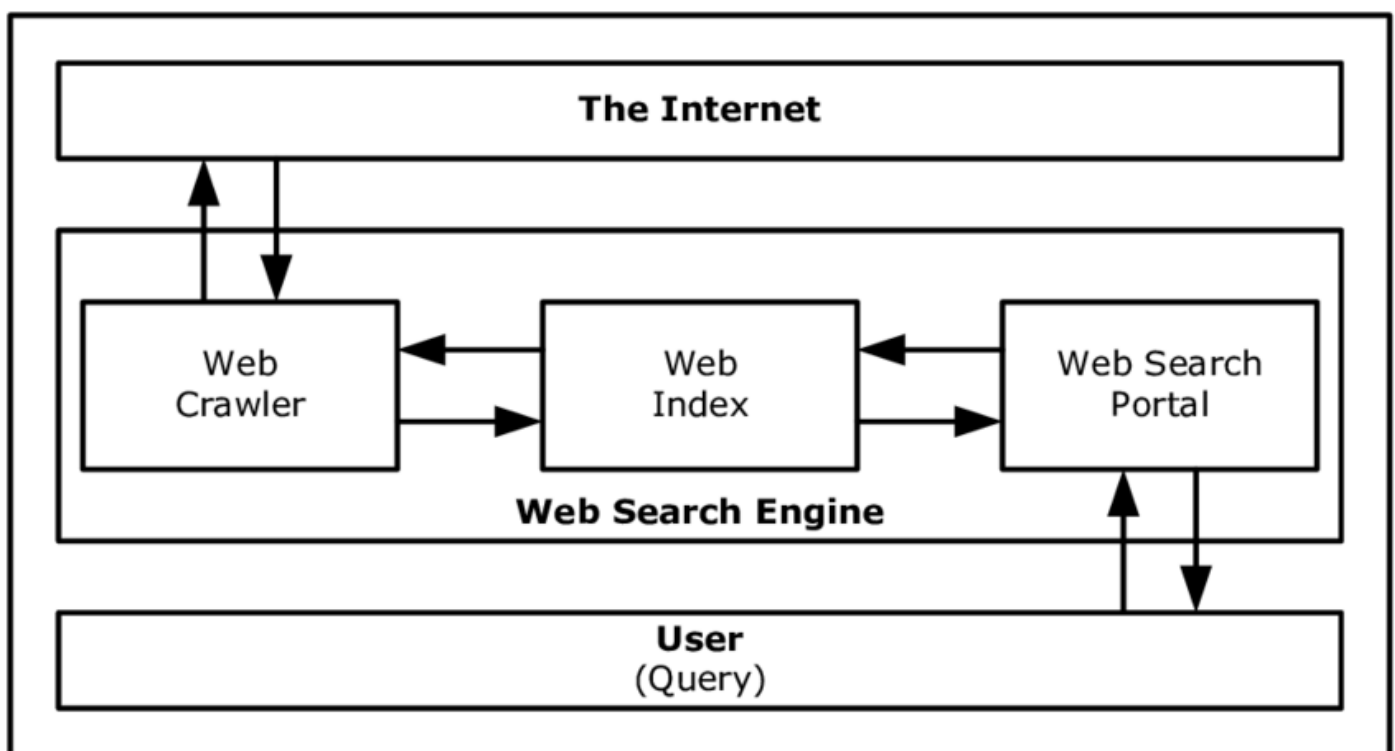


Fig. 5. Arquitectura de un motor de búsqueda en internet ([fuente original](#))

Cuando un *robot* localiza una nueva página o documento, procede a enviar una copia del mismo, o de partes seleccionadas, al programa de representación e indización que utiliza el motor de búsqueda, y que es el encargado de generar una representación de la página o documento para su incorporación a la base de datos. Sus características permiten que los responsables de la programación de los robots establezcan parámetros para su actuación, atendiendo a factores como tamaño del web que lo contiene, tiempo desde la última actualización en la base de datos, “popularidad” de los documentos en las búsquedas de los usuarios, etc.

Aunque no existe una norma estándar para la arquitectura y actividad de los robots, hay que considerar que en 1993 y 1994 ya se recomendó un estándar de facto, que no ha sido aplicado en

su totalidad. Las Guías para Autores de Robots y las [Normas para la Exclusión de Robots](#) establecen los principios de actividad responsable, eficiencia en la utilización de recursos y limitaciones de acceso a los servidores de información. Estos principios establecen la posibilidad de que los responsables de un sitio web excluyan al sitio en cuestión de ser recopilado y, en consecuencia, de ser incluidos sus contenidos en los índices que los motores de búsqueda usan para las tareas de recuperación de información.

Material complementario

- [Motores de búsqueda: qué son, cómo funcionan y cuántos tipos existen \(2020\)](#)
- [¿Cómo Funcionan los Motores de Búsqueda? Una Guía Para Principiantes \(2021\)](#)

Financiado por el Ministerio de Educación y Formación Profesional y por la Unión Europea - NextGenerationEU



2.2. Procesamiento y recuperación de información

Una vez que el motor recibe la copia del documento que le remite el robot, se procede a su representación o indización. Esto es debido a que las páginas o documentos no pueden integrarse directamente en las bases de datos. Previamente **es necesario llevar a cabo un proceso de indización automática, con la finalidad de crear una representación del contenido informativo**. Esta indización puede realizarse desde diferentes enfoques, aunque **el más utilizado es el modelo que se basa en la representación de los documentos como vectores, dentro de un espacio vectorial**.

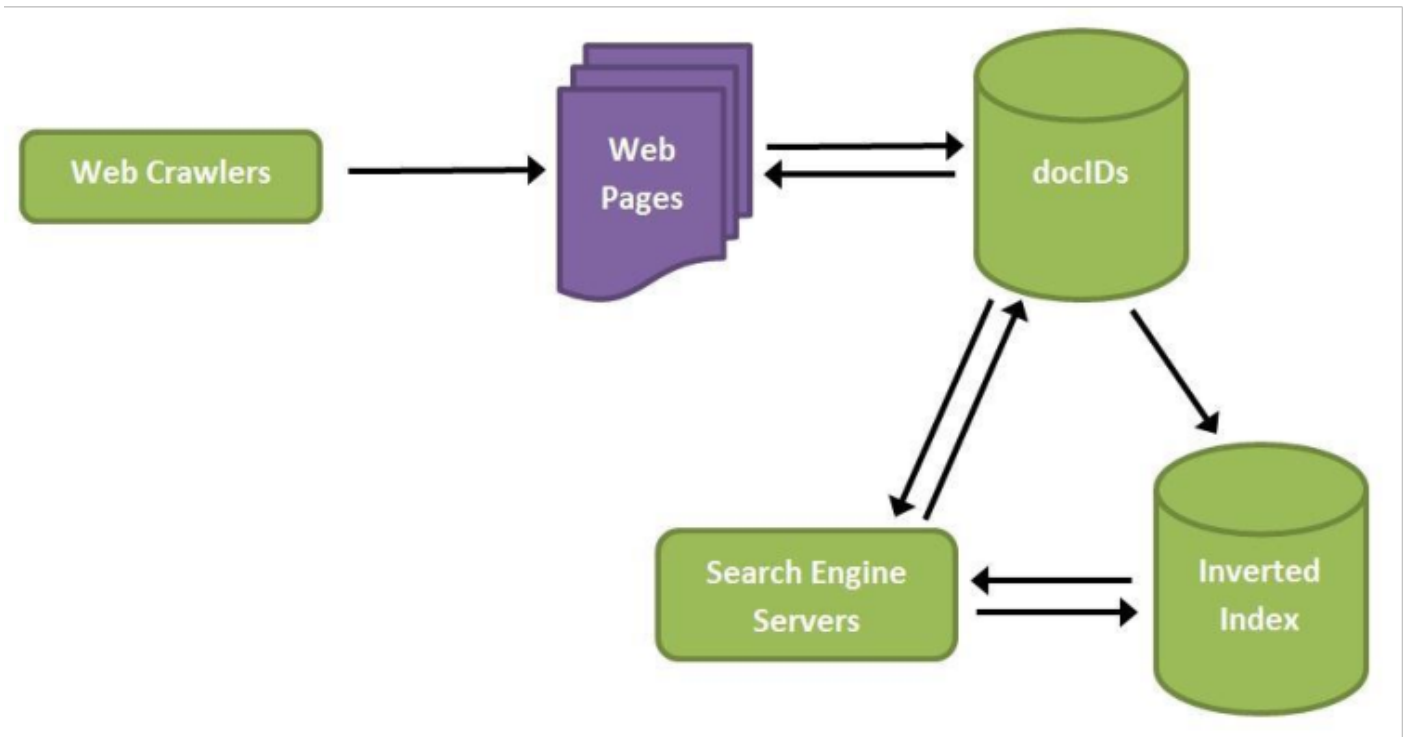


Fig. 6. Esquema de funcionamiento del proceso de captura y representación en un motor de búsqueda ([fuente original](#)).

Toda esta información se almacena en grandes bases de datos, que además pueden incorporar a cada representación otros datos, como el url ([Universal/Uniform Resource Locator](#)) de cada página o documento, o dar prioridad en los vectores a algunos elementos específicos, como puedan ser los metadatos detectados.



Financiado por
la Unión Europea
NextGenerationEU



Plan de Recuperación,
Transformación
y Resiliencia



GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN
Y FORMACIÓN PROFESIONAL



GOBIERNO
DE ARAGON

2.3. Presentación de resultados

La explotación del contenido de los motores se lleva a cabo mediante la consulta a través de una interfaz web. Los motores suelen ofrecer una interfaz sencilla o simplificada, que presenta la típica caja de diálogo para introducir una expresión, y una interfaz avanzada, que ofrece más opciones para establecer parámetros que refinan y limitan los ámbitos de búsqueda. Estas interfaces permiten enviar al motor de recuperación la expresión o ecuación de búsqueda formulada por el usuario. Este componente, sobre el que los motores son poco dados a ofrecer información, es la clave para la recuperación de información, ya que es el responsable de traducir la ecuación de búsqueda, en formato textual, a una expresión matemática que es la confrontada con el contenido de la base de datos.

La presentación de las respuestas obtenidas se basa en la presentación de listados de respuestas, ordenados de mayor a menor relevancia según cálculo propio del motor de recuperación. El conocido [PageRank](#) de Google es un algoritmo de ordenación que utiliza diferentes criterios para asignar la relevancia. Estas listas, generadas de forma dinámica como resultado del proceso de interrogación, contienen el listado de respuestas, indicando para cada una de ellas su título, url, fecha, tamaño, un breve resumen que suele corresponder a las primeras líneas o a las cabeceras del contenido, y otros datos. Los resultados se muestran ordenados según la pertinencia dada a cada documento, comparándolo con la expresión formulada por el usuario. En los motores también se incorporan opciones que ofrecen acceso a resultados similares o del mismo tipo, utilizando para ello precisamente la distancia entre los vectores que representan documentos dentro del espacio vectorial. Finalmente, tras el listado de respuestas se ofrece de nuevo la interfaz de interrogación, para que el usuario pueda reformular la expresión de búsqueda si es necesario.

Los motores de búsqueda, excepto algunas excepciones, son un negocio. Están sostenidos y operados por empresas comerciales, lo que supone que junto a los resultados de una búsqueda es posible encontrar enlaces que no son resultados de esa búsqueda. Hay que diferenciar entonces entre lo que se llama “[búsqueda orgánica](#)”, y lo que no. **Búsqueda orgánica es el conjunto de resultados que no son anuncios, y que proceden de un proceso de búsqueda y presentación neutral.** En los listados de respuestas, aquellos resultados que no son orgánicos, que son publicidad, deben presentarse siempre identificados como “Anuncio”.



Hembrillas, escarpías y alcayatas - Leroy Merlin



Te ofrecemos una amplia colección de hembrillas, **escarpías** y alcayatas para todo tipo materiales.

<https://www.amazon.es> > escarpías > k=escarpías ▾

1-48 de 706 resultados para "escarpías" - Amazon.es



Amazon.es: **escarpías**. ... Ganchos de Tornillo, 100 Pcs Alcayata en Forma de L, Pequeños Alcayatas Pared de Madera, L **Escarpías** Tornillería Cancamos para ...

<https://www.bricomart.es> > clavos-hembrillas-y-alcayatas ▾

Escarpías | Bricomart

Descubre nuestra amplia gama de **escarpías**. Las mejores marcas y la mejor calidad, en Bricomart.

<https://www.bauhaus.es> > pernos-hembrillas-y-escarpías

Pernos, hembrillas y escarpías | BAUHAUS



Escarpia roscada. Ø x L: 3,8 x 15 mm, 8 ud., Rosca de madera, Baño de latón, Con saliente. 2

Variantes. 3,19 €. por unidad. Disponible online.

Fig. 7.

Una búsqueda

poco orgánica ([fuente original](#)).

Progresivamente, los motores han ido añadiendo más prestaciones de cara a los usuarios, como el mantenimiento del historial de búsqueda, el establecimiento de alertas personalizadas con nuevos resultados sobre una búsqueda dada, o la posibilidad de compartir resultados con otros usuarios.

Financiado por el Ministerio de Educación y Formación Profesional y por la Unión Europea - NextGenerationEU

