

## 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.

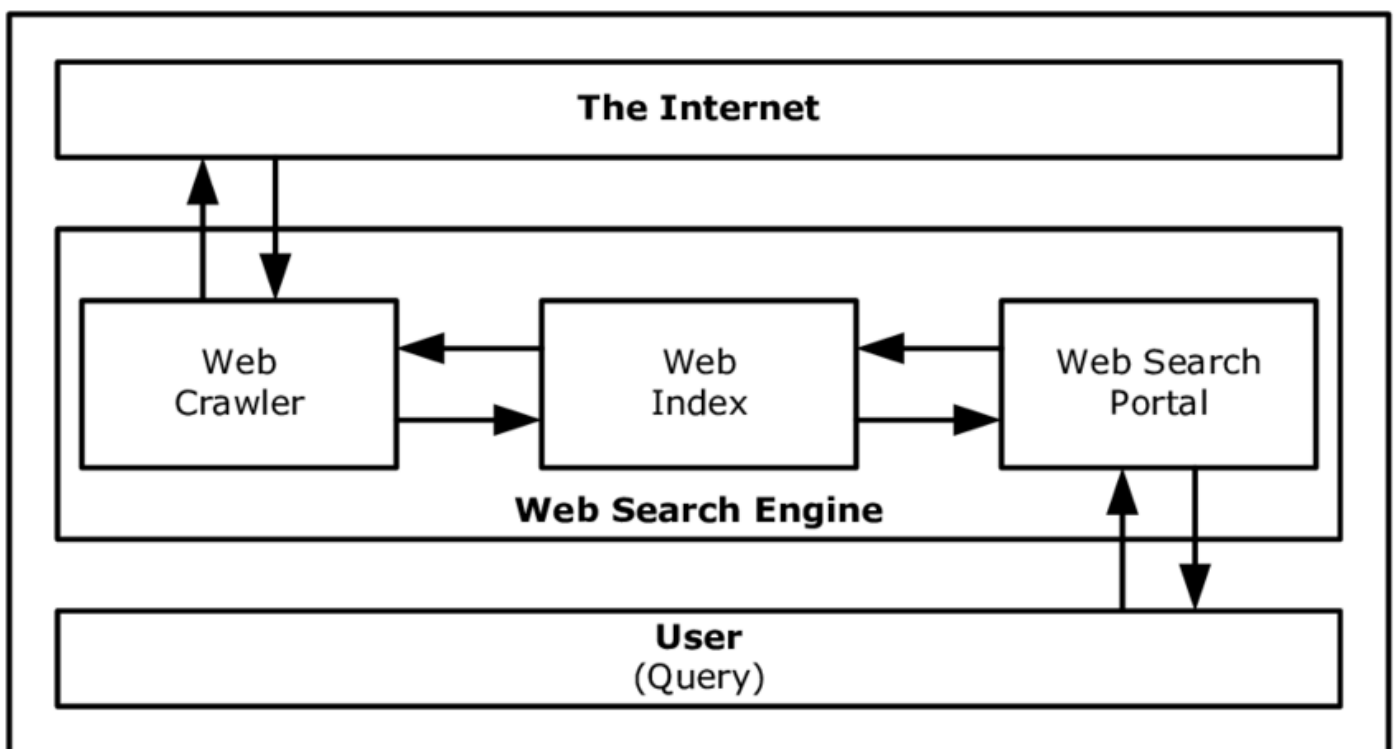


Fig. 5. Arquitectura de un motor de búsqueda en internet (fuelle original)

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



Revision #7

Created 27 October 2022 18:27:37 by Jesús Tramullas

Updated 18 January 2023 09:44:58 by Equipo CATEDU