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**.

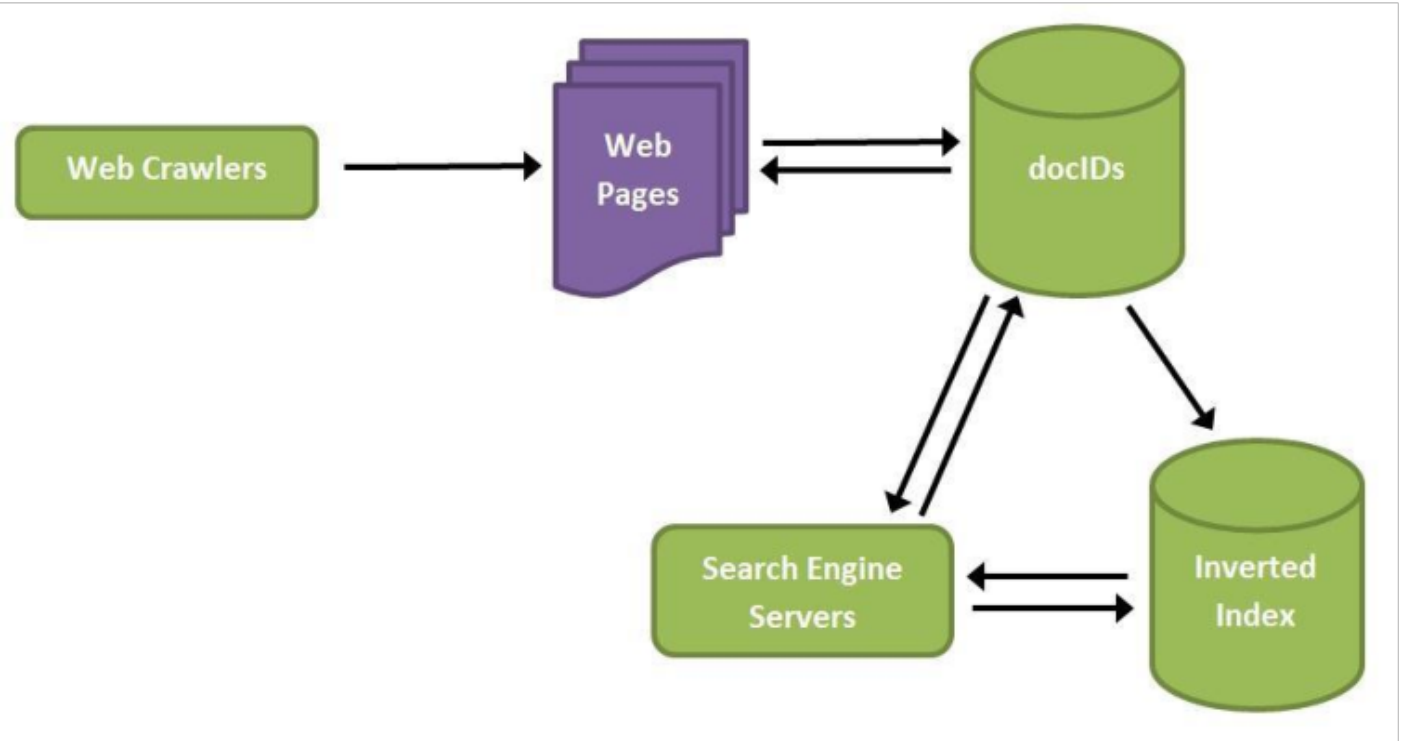


Fig. 6. Esquema de funcionamiento del proceso de captura y representación en un motor de búsqueda (fuente original).

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.

Copyright 2026

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



Revision #7
Created 2022-10-27 18:27:56 CEST by Jesús Tramullas
Updated 2023-04-08 19:26:54 CEST by Ana López Floría

Copyright 2026