

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**.

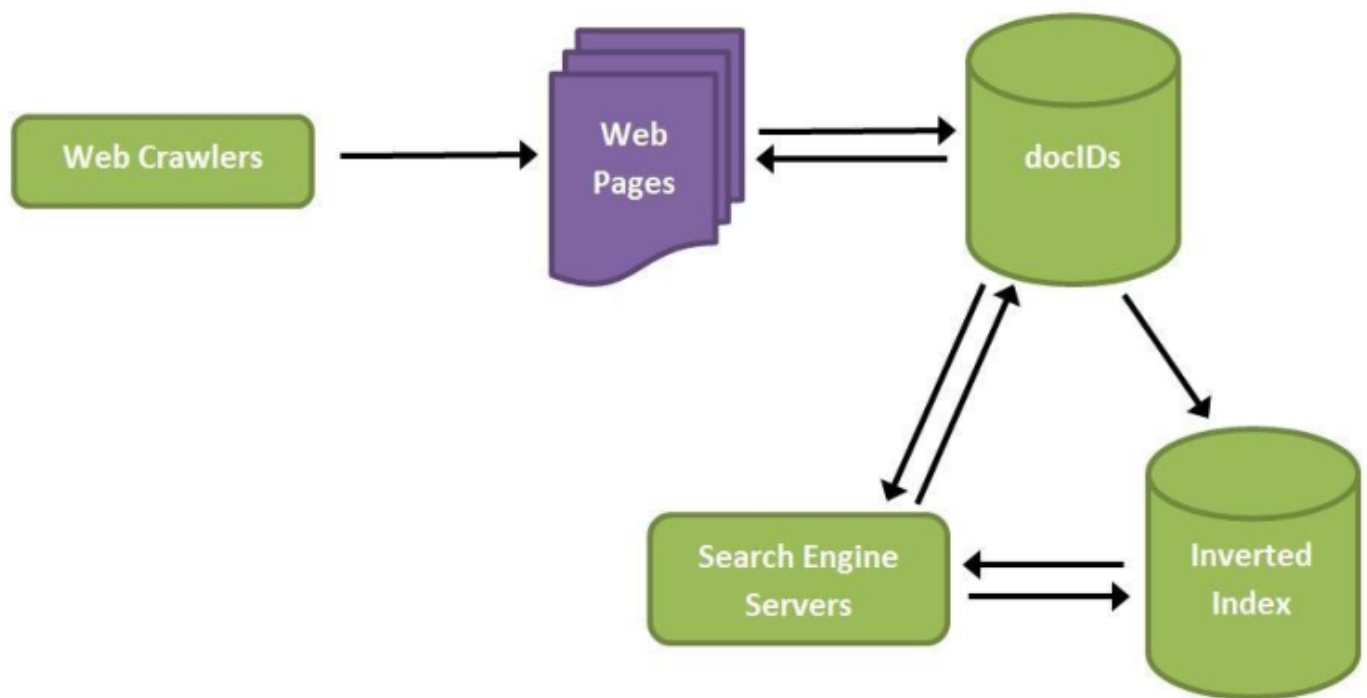


Fig. 6. Esquema de funcionamiento del proceso de captura y representación en un motor de búsqueda ([fuente original](#)).

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



Financiado por
la Unión Europea
NextGenerationEU



Plan de Recuperación,
Transformación
y Resiliencia



GOBIERNO
DE ARAGON



GOBIERNO
DE ARAGON

Revision #7

Created 27 October 2022 18:27:56 by Jesús Tramullas

Updated 8 April 2023 19:26:54 by Ana López Floría