

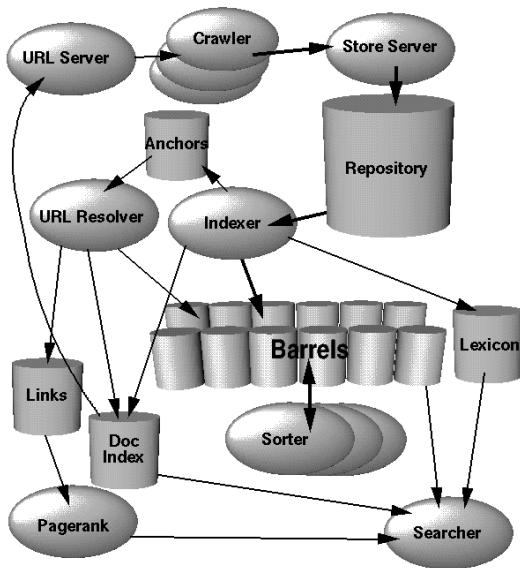
Ingeniería de sistemas de información

Búsquedas en Internet

Daniel López García
Lothar Soto Palma
Universidad de Granada

3 de mayo de 2017

Arquitectura de Google



- Las estructuras de datos usadas están optimizadas para manejar grandes colecciones de datos.
- Se hace uso del modelo MapReduce para la creación de los índices.
- El crawling se realiza usando una gran cantidad de crawlers de forma distribuida.
- Las páginas webs obtenidas mediante el crawling se almacenan comprimidas en un repositorio. La función de indexación descomprime los documentos y los parsea para crear el índice.
- Otro aspecto a tener en cuenta es el texto de los enlaces, ya que normalmente contienen la mejor descripción de la página.

Proporciona una medida de la importancia de la página en función de las páginas que contienen enlaces a la misma. Este valor se calcula sumando el número de enlaces a la pagina normalizados por un parámetro d que indica la importancia de la página de la que procede el enlace.

$$PR(A) = (1 - d) + d\left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)}\right)$$