Abstract

The Internet has become a cosmic information source in recent years and can be considered as the world's largest digital library. To aid ordinary users in finding desired data in this library, numerous search engines have been created. Each search engine has a corresponding database that defines the set of documents that can be searched by the search engine.
Typically, an index for all documents in the database is created and stored in the search engine. Text data in the Internet can be partitioned into numerous databases naturally. Proficient retrieval of desired data can be realized if we can accurately envisage the usefulness of each database, because with such information, we only need to retrieve potentially useful documents from useful databases. For a given query ‘q’ the usefulness of a text database is defined to be the no. of documents in the database that are sufficiently relevant to the query ‘q’.

In this paper, we propose innovative approaches for database selection and documents selection.

Reference


Index Terms

Computer Science

Information Retrieval

Key words

Metasearch Engine

Distributed query processing

Document selection