A Novel Technique for Database Selection and Document Selection

International Journal of Computer Applications
© 2011 by IJCA Journal

Number 8 - Article 4

Year of Publication: 2011

Authors:

Anil Agrawal
Mohd. Husain
Raj Gaurang Tiwari
Subodh Kumar

10.5120/2241-2865
{bibtex}pxc3872865.bib{/bibtex}

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

- l. gravano and h. garcia-molina, "generalizing gloss to vector-space databases and broker hierarchies," int'l conf. very large data bases, p. 78-89, sep. 1995.
- c. badue, r. baeza-yates, b. ribeiro-neto, and n. ziviani. distributed query processing using partitioned inverted files. in proc. of the 9th string processing and information retrieval symposium (spire), september 2002.
- boris chidlovskii, claudia roncancio, and marie-luise schneider. semantic cache mechanism for heterogeneous web querying. in proceedings of the www8 conference / searching and querying, 1999.

Index Terms

- Computer Science
- Information Retrieval

Key words

- Metasearch Engine
- Distributed query processing
- Document selection