Designing New Crawling and Indexing Techniques for Web Search Engines

By Tan, Qingzhao

Condition: New. Publisher/Verlag: VDM Verlag Dr. Müller | A New Generation of Web Search Engine | This thesis studies in a Web search engine how a crawler with limited computing resource can effectively crawl from the dynamically changing Web and acquire the most updated Web documents, and how a Web search engine can provide information-object--oriented indexing methods which enable users to retrieve desired information with high accuracy and high efficiency. To address the first problem, we design a set of sampling policies with various downloading granularity for the sampling method, taking into account the link structure, the directory structure, and the content-based features which include the clustering technique. We further extend the clustering-based sampling approach by testing more dynamic features and strategically selecting samples from each cluster. For the second problem, we propose building indexes on extracted metadata of various information objects, instead of the whole document. We set up a digital library named ArchSeer for the domain of archeology. ArchSeer allows users to retrieve archeology literature via domain-specific search engines. | Format: Paperback | Language/Sprache: english | 224 gr | 220x150x8 mm | 156 pp.

READ ONLINE
[6.09 MB]

Reviews

Undoubtedly, this is actually the finest work by any writer. It is really basic but excitement within the fifty percent of your publication. Your way of life period is going to be enhance as soon as you comprehensive looking over this ebook.
-- Matt Maggio

The book is fantastic and great. It is filled with wisdom and knowledge I am just easily will get a enjoyment of looking at a composed publication.
-- Bradley Hahn