1. Design and implementation of a distributed high-performance web crawler

Accession number: 2004278248949

Authors: Zhang, Ling (1); Ye, Yun-Ming (1); Song, Hui (1); Yu, Shui (1); Ma, Fan-Yuan (1)
Author affiliation: (1) Dept. of Comp. Sci. and Eng., Shanghai Jiaotong Univ., Shanghai 200030, China
Corresponding author: Ma, F.-Y.(fyma@sjtu.cn)

Source title: Shanghai Jiaotong Daxue Xuebao/Journal of Shanghai Jiaotong University
Abbreviated source title: Shanghai Jiaotong Daxue Xuebao
Volume: 38
Issue: 1
Issue date: January 2004
Publication year: 2004
Pages: 59-61
Language: Chinese
ISSN: 10062467
CODEN: SCTPDH
Document type: Journal article (JA)
Publisher: Shanghai Jiao Tong University

Abstract: Web crawler is the core component of WWW search engine and information retrieval systems. This paper discussed the architecture of a distributed Web crawler and the design ideas about the Web crawler data structure, system modules and related algorithms. The key problems encountered in the design and implementations were also commented, and the solutions to those problems were presented.

Number of references: 5
Main heading: Search engines
Controlled terms: Distributed parameter networks - Information analysis - Information retrieval - Java programming language - World Wide Web
Uncontrolled terms: Online search - Web crawler
Classification code: 723.1.1 Computer Programming Languages - 903.1 Information Sources and Analysis - 921 Mathematics
Treatment: Applications (APP)
Database: Compendex

Data Provider: Engineering Village