Abstract

Actuality

The number of information sources increases with each passing day, the number of accumulated data is growing. In the avalanche of information work is almost impossible for human being (a human being cannot respond to the information changes). While reading news man filters the information he needs and gets rid of other information not even looking through the text, but focusing only on the news rating or its name. The teacher cannot read through every work of each of his students. It becomes impossible to navigate the mass of books.

The purpose and tasks of research

The aim of this work is to examine aspects of information retrieval systems on a given set of sites, as technologies for the collection of information about the contents of Web resources.

Problems solved during research

The main objectives of the developed software, designed to estimate the parameters of the web pages are:

1. review the organizational structure of sites network NTU "KPI";
2. development of automated data collection module;
3. review existing models of information retrieval;
4. review the organization of the Semantic Web;
5. review of existing solutions consider solutions for the full-text search;
6. develop search engine on a given set of sites, including all phases;
7. achieve high performance results

Results achieved

The result of the work is a web-based application designed to control the information placed on web pages portals NTU "KPI" and EGEE, in order to improve the rating NTU "KPI" in the world ranking universities in Webometrics.

Scientific novelty of the results

Scientific novelty of the work is to study the possibility of using web-based applications for searching information displayed on the web.

The practical significance of the results

The practical application of the developed software is the ability to use it to find the information provided on websites NTUU "KPI", thereby simplifying the process find specific information.

Findings

In this master's work various systems designed for find information search were investigated. To implement the data retrieval in the Web space NTUU "KPI" a search module was developed and implemented. This module is a part of the integrated system to control the information provided on the web.

During the work the following tasks were solved:

1. The choice of indexing and search mechanism that satisfies the relation of efficiency and speed of search and computational capabilities that are needed for this.

2. Implementation of the full-text search on the information collected from specified websites.

3. Achieving a high speed information processing.

4. Building a user interface for search.

The experiments showed that the quality of search engines is sufficient to use the results of its work in dealing with future applications.

Work on the 81 pages contains 4 tables, 3 illustrations. During the work literature from 21 different sources was used.

Keywords: search, robot, crawler, web, web portal, an index model.