PROCEEDINGS

of the Union of Scientists - Ruse

Book 5

Mathematics, Informatics and Physics

Volume 11, 2014

RUSE

PROCEEDINGS
OF THE UNION OF SCIENTISTS - RUSE

Union of Scientists - Ruse
16, Konstantin Irechek Street
7000 Ruse
BULGARIA
Phone: (++359 82) 828 135,
(++359 82) 841 634
E-mail: suruse@uni-ruse.bg
web: suruse.uni-ruse.bg

Contacts with Editor
Phone: (++359 82) 888 738
E-mail: zzdravkova@uni-ruse.bg

PROCEEDINGS
of the Union of Scientists – Ruse

ISSN 1314-3077

ISSN 1314-3077

Proceedings

of the Union of Scientists– Ruse

Contains five books:

1. Technical Sciences
2. Medicine and Ecology
3. Agrarian and Veterinary Medical Sciences
4. Social Sciences
5. Mathematics, Informatics and Physics

The Ruse Branch of the Union of Scientists in Bulgariawas foundedin 1956. Its first Chairman was Prof. StoyanPetrov. He was followed by Prof. TrifonGeorgiev, Prof. KolyoVasilev, Prof. Georgi Popov, Prof. MityoKanev, Assoc. Prof. Boris Borisov, Prof. Emil Marinov, Prof. HristoBeloev. The individual members number nearly 300 recognized scientists from Ruse, organized in 13 scientific sections. There are several collective members too– organizations and companies from Ruse, known for their success in the field of science and higher education, or their applied research activities. The activities of the Union of Scientists – Ruse are numerous: scientific, educational and other humanitarian events directly related to hot issues in the development of Ruse region, including its infrastructure, environment, history and future development; commitment to the development of the scientific organizations in Ruse, the professional development and growth of the scientists and the protection of their individual rights.

The Union of Scientists – Ruse (US – Ruse) organizes publishing of scientific and popular informative literature, and since 1998 – the "Proceedings of the Union of Scientists- Ruse".

BOOK 5

"MATHEMATICS, INFORMATICS AND PHYSICS"

VOLUME 11

CONTENTS

Mathematics

Informatics

BOOK 5

"MATHEMATICS, INFORMATICS AND PHYSICS"

VOLUME 11

Physics

web: suruse.uni-ruse.bg

UPDATING THE RECORDS OF THE SEARCH ENGINES DUE TO A CLIENT REQUEST

Metodi Dimitrov

Angel Kanchev University of Ruse

Abstract: *Nowadays, the search engines are very commonly used. To be useful, however, they must maintain good enough database and must possess good enough algorithms for searching and evaluation so that they will be able to find and serve the desired websites. To complete those tasks, the search engines must possess a sufficiently large number of computers and computing power and must periodically crawl the Internet. This centralization of resources makes the work of the search engine not very productive.*

The current paper proposes a new method of updating the records of the search engines. The proposed method reduces the unnecessary work, carried out by the search engines and helps them to eliminate the obsolescence in their database.

Keywords: *search engine, records, client request*

INTRODUCTION

Development of the Internet has led to the appearance of trillion web pages [3] with different content. Because of that large number of pages and that great diversity, when a consumer needs to access certain information, he has no choice but to turn to the search engine. To be competitive, of course, the results should be adequate to customer demand. This means that the search engines must:

1. Crawl and periodically be crawling enough web addresses to accumulate a large enough database, which covers the diverse information needs of the users;
2. Have enough good algorithms for searching and evaluation so that they will be able to find and serve the websites, in which the customers would be interested most.

After analyzing these criteria, it can be said that in order to be covered, different things are required. While the first criterion needs an accumulation (possession) of a sufficiently large number of computers and computing power for crawling the web, the second criterion needs more intellectual work for developing certain search rules and software.

If we focus on the first criterion, we can deduce the following statements:

- Crawling on all web pages (addresses) is impossible, because at any moment new URL addresses are being registered. At the same time, for extending its databases, search engines rely on the hyperlinks following (from one page to another). In this context, there will always be isolated and closed groups of URL addresses that remain invisible to search engines;
- Crawling and analyzing the content of most of the web addresses on the Internet
- needs a large number of computers and huge computing power;
- Even if a given search engine checks the content of a web page and update its registry, it may be that the content of the registry is outdated, because soon after the crawling, the administrators of the website could replace its content.

The current report offers a solution to the above problems associated with the maintenance of the correct content in the databases of the search engines by proposing a new approach for update of the search engines' registers.

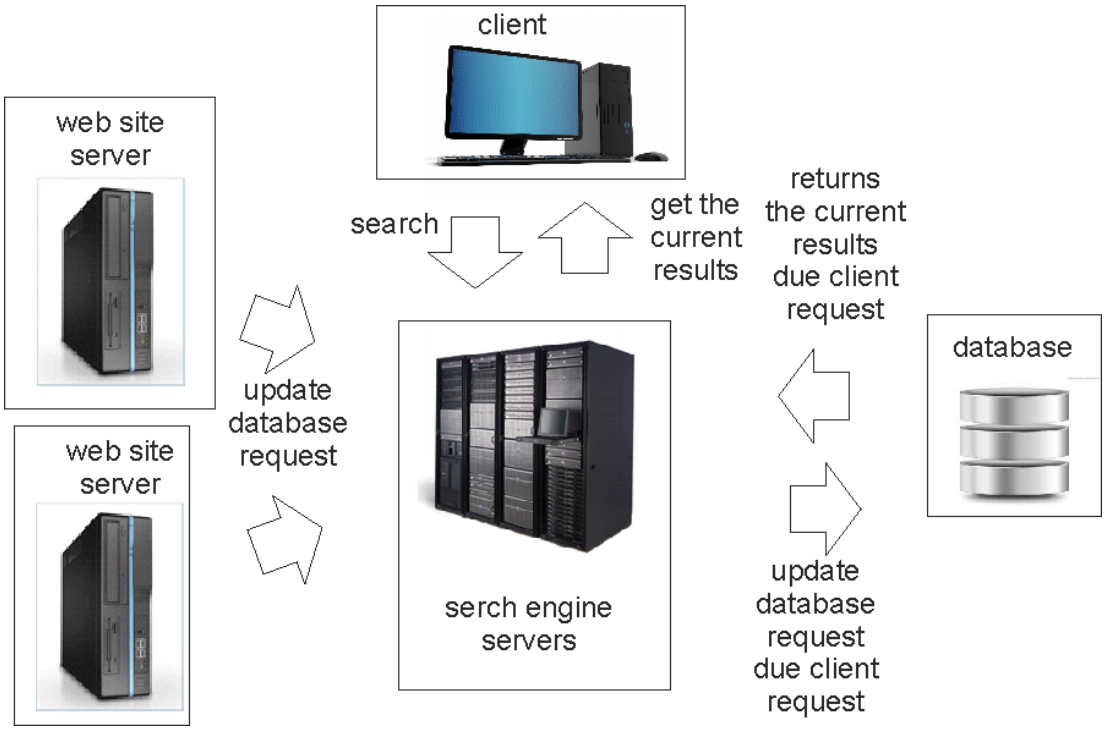

Fig. 1. Updating the database of the search engines due to a client request.

UPDATING THE RECORDS OF THE SEARCH ENGINES DUE TO A CLIENT REQUEST

Currently, the update of the search engines' databases is performed by crawling through the websites (URL addresses). As mentioned above, this is actually an inefficient process. As an alternative to the status quo a new, alternative method is proposed. According to the method, every time the content of a web page is changed, it notifies the search engines for that change. Thus, immediately after the update of a given web-page content, search engine updates its database. So search engines always have the latest information for every web page/site. Fig. 1 shows the scheme of updating the search engine's database due to the client request.

The proposed approach will also reduce the need of concentration of funds and resources on one place, by distributing the work on updating of the database of the search engines. Another advantage of using the proposed method is that it will prevent the performance of the work in vain, performed by the search engines. Currently, the content of a given website can be unchanged for years, but to keep their database up to date, the search engines still need to crawl it periodically.

PROTOCOL FOR UPDATING THE RECORDS OF THE SEARCH ENGINES DUE TO A CLIENT REQUEST

The protocol aim is to establish the rules by which the clients (web sites) send information to the search engines. The protocol is JSON based and transmits the following data:

- The URL address of the web page.
- The content of the web page.

- The date, on which the web page contained the sent content.

Fig. 2 shows an example of the JSON structure which transmits data to the search engines.

```
{
        "url":"the-ulr-address-goes-here",
        "data":"the-web-page-code-goes-here ",
        "date":"the-date-goes-here "
}
```

Fig. 2. Example of the protocol, which transmits data to the search engines.

LIBRARY FOR UPDATING THE RECORDS OF THE SEARCH ENGINES DUE TO A CLIENT REQUEST

PHP based library named KMP (Keep Me Posted) is developed. Its purpose is to automate the process of refreshing the database of the search engines, after the given web page changes its content. The library is working with its own database (table), where different versions of the web page are saved.

Activation of the library is done with inclusion in any web page, which must notify the search engine in case of change of its content. Fig. 3 shows an example of the library usage.

After including the library on certain web page, if it detects a change in the content, notifies the search engine, so it can update its database.

```
<?php
        //KMP Library Start

        include("./kmp/the_library.php");

        //KMP Library End
?>
```

Fig. 3. Library usage in a web page.

CONCLUSIONS AND FUTURE WORK

The following conclusions can be made:

1. Approach for updating the records of the search engines due to a client request is proposed. The approach reduces unnecessary work carried out by the search engines and helps to eliminate obsolescence in their database.

2. Protocol by which the search engines update their database is developed. The protocol is based on JSON.

3. Library that monitors the change in the content of a given web page and sends information to the search engines, if detecting such, is developed.

4. The developed library sends the contents of the web page in which it is included to a web server (URL address or a search engine). Therefore, the library can be used for

storing the history of the content of the web page. In this sense, the system can be developed, by being integrated into a system for backing up the content of web pages.

The present document has been produced with the financial assistance of the European Social Fund under Operational Programme "Human Resources Development". The contents of this document are the sole responsibility of "Angel Kanchev" University of Ruse and can under no circumstances be regarded as reflecting the position of the European Union or the Ministry of Education and Science of Republic of Bulgaria.

Project № BG051PO001-3.3.06-0008 "Supporting Academic Development of Scientific Personnel in Engineering and Information Science and Technologies"

REFERENCES

[1] Buettcher, S., C. Clarke, G. Cormack, Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press, 2010.

[2] Croft, B., D. Metzler T. Strohman, Search Engines: Information Retrieval in Practice, Addison-Wesley, 2009.

[3] Kelly, K., What Technology Wants, Penguin Group US, 2010.

CONTACT ADDRESS

Pr. Assist. Metodi Dimitrov, PhD
Department of Informatics and Information Technologies
Faculty of Natural Sciences Education
Angel Kanchev University of Ruse
8 Studentska Str., 7017 Ruse, Bulgaria
Phone: (++359 82) 888 470
E-mail: mdimitrov@ami.uni-ruse.bg

ОБНОВЯВАНЕ НА РЕГИСТРИТЕ НА ТЪРСЕЩИТЕ МАШИНИ ПО ИСКАНЕ НА КЛИЕНТА

Методи Димитров

Русенски университет „Ангел Кънчев"

Резюме: *В днешно време, търсещите машини са много често използвани. За да бъдат полезни обаче, те трябва да притежават достатъчно добра база данни и достатъчно добри алгоритми за търсене и оценка така, че да могат да намерят и поднесат подходящата информация. За да изпълнят тази задача, търсещите машини трябва да притежават голяма изчислителна мощ и периодично да обхождат уеб пространството. Тази централизация на ресурси обаче, прави търсещите машини не много продуктивни.*

Текущия доклад предлага нов метод на обновяване на регистрите на търсещите машини. Предложеният подход намалява количеството ненужна работа, извършваното от търсещите машини и им помага да премахнат неточностите в базата им данни.

Ключови думи: *търсеща машина, регистри, клиентска заявка*

Requirements and guidelines for the authors - "Proceedings of the Union of Scientists - Ruse" Book 5 Mathematics, Informatics and Physics

The Editorial Board accepts for publication annually both scientific, applied research and methodology papers, as well as announcements, reviews, information materials, adds. No honoraria are paid.

The paper scripts submitted to the Board should answer the following requirements:

1. Papers submitted in English are accepted. Their volume should not exceed 8 pages, formatted following the requirements, including reference, tables, figures and abstract.

2. The text should be computer generated (MS Word 2003 for Windows or higher versions) and printed in one copy, possibly on laser printer and on one side of the page. Together with the printed copy the author should submit a disk (or send an e-mail copy to: vkr@ami.uni-ruse.bg).

3. Compulsory requirements on formatting:

- font - Ariel 12;
- paper Size - A4;
- page Setup - Top: 20 mm, Bottom: 15 mm, Left: 20 mm, Right: 20mm;
- Format/Paragraph/Line spacing - Single;
- Format/Paragraph/Special: First Line, By: 1 cm;
- *Leave a blank line under Header - Font Size 14;*
- Title should be short, no abbreviations, no formulas or special symbols - Font Size 14, centered, Bold, All Caps;
- *One blank line - Font Size 14;*
- Name and surname of author(s) - Font Size: 12, centered, Bold;
- *One blank line - Font Size 12*;
- Name of place of work - Font Size: 12, centered;
- *One blank line;*
- abstract – no formulas - Font Size 10, Italic, 5-6 lines ;
- keywords - Font Size 10, Italic, 1-2 lines;
- *one blank line;*
- text - Font Size 12, Justify;
- references;
- contact address - three names of the author(s) scientific title and degree, place of work, telephone number, Email - in the language of the paper.

4. At the end of the paper the authors should write:

- The title of the paper;
- Name and surname of the author(s);
- abstract; keywords.

Note: The parts in item 4 should be in Bulgarian and have to be formatted as in the beginning of the paper.

5. All mathematical signs and other special symbols should be written clearly and legibly so as to avoid ambiguity when read. All formulas, cited in the text, should be numbered on the right.

6. Figures (black and white), made with some of the widespread software, should be integrated in the text.

7. Tables should have numbers and titles above them, centered right.

8. Reference sources cited in the text should be marked by a number in square brackets.

9. Only titles cited in the text should be included in the references, their numbers put in square brackets. The reference items should be arranged in alphabetical order, using the surname of the first author, and written following the standard. If the main text is in Bulgarian or Russian, the titles in Cyrillic come before those in Latin. If the main text is in English, the titles in Latin come before those in Cyrillic. The paper cited should have: for the first author – surname and first name initial; for the second and other authors – first name initial and surname; title of the paper; name of the publishing source; number of volume (in Arabic figures); year; first and last page number of the paper. For a book cited the following must be marked: author(s) – surname and initials, title, city, publishing house, year of publication.

10. **The author(s) and the reviewer, chosen by the Editorial Board, are responsible for the contents of the materials submitted**.

Important for readers, companies and organizations

1. Authors, who are not members of the Union of Scientists - Ruse, should pay for publishing of materials.

2. Advertising and information materials of group members of the Union of Scientists – Ruse are published free of charge.

3. Advertising and information materials of companies and organizations are charged on negotiable (current) prices.

Editorial Board

ISSN 1314-3077