

Possible Methods of Improving the Quality of Search and Ranking Information on the Internet

Viacheslav Zosimov

Mykolayiv V.O. Sukhomlynsky National University, 54030, Ukraine, Mykolayiv, Nikolska str., 24

zosimovvv@bk.ru

Abstract. *The paper describes main reasons for the poor quality of the Internet search. And Possible ways to improve the quality of information search on the Internet.*

Keywords

Inductive modeling, GMDH, information search, relevance, ranking.

Introduction

Search results ranking methods of modern search engines are no longer cope with their task.

Almost any website can be put on the first search results page by artificially raising the necessary ranking parameters. These options are no longer a secret, and they can be increased without much material costs.

An alternative to the modern search engines are search agents. They allow very high efficiency for specific information search. But, usually configured on certain types of tasks and conduct search in a limited number of documents.

Therefore, developing of new universal web search and ranking methods is an actual task

The main reasons for the poor quality of the Internet search

1. The large number of sites with the "rewrite" content. The same text is rewritten in other words, several tens or even hundreds times by the owners of various sites and placed to attract customers. This is the easiest and most inexpensive way to get a new unique content. Therefore, it is most in demand. For certain categories of information, such clones are up to 90% of the first few search results pages. Most often this occurs on the news search. As a result, from the entire set of documents relevant to the query, user gets only a few alternatives and its clones

2. It is simple to cheat search engines ranking system. The easiest way - to buy external links to web-site on "links stock". There are a number of web resources that allow for a fee to place external links on the pre-recorded advertising sites. Practice shows that current search engines do not track such links. Fig. 1 shows a plot of the position of the site in search results before and after the purchase of external links.

3. It does not take into account the context in which the information being sought.

4. There is no way to choose a category of information search.

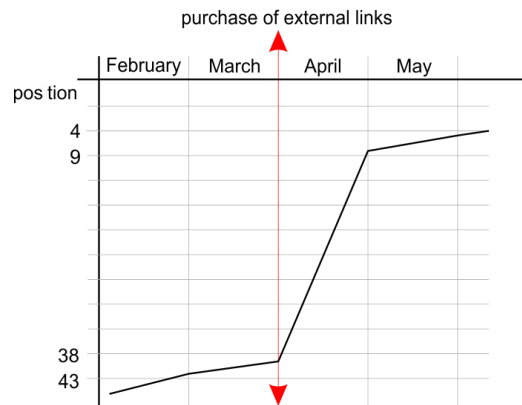


Fig 1. Schedule of site position changes

Possible ways to improve the quality of information search on the Internet

1. It is necessary to develop methods to detect not just a clone of the original document but also its various modifications. This raises two further problems:

- detection of the original document;
- setting the sensitivity of the classifier that will decide what counts as a clone, and that a unique document different from the original.

2. It is necessary to develop methods for detection the external links posted through exchanges.

It is necessary to develop a new websites ranking system fundamentally different from modern search engines models. It need to identify the signs, that are most distinctly indicates web resource quality.

To build the ranking model on the basis of received signs it should be used GIA GMDH, which has perfectly proved itself in previous experiments on the construction of search engines Google and Yandex ranking models [2].

3. The development of methods based on the search query analysis, to determine the context in which it has been used.

4. The division of information into several categories to narrow search area, and the issuance of the most relevant results.

Conclusion

Solution of the described problems requires the development of a new search engine, which at the stage of data collecting from website will collect all the necessary information for further analysis. The new system should be used ranking methods based on attributes that are difficult to falsify. It is also necessary to develop methods of linguistic and semantic analysis of the sites content to identify clones and original documents. This approach involves the use of cloud technologies and parallel computing for storing and processing large amounts of data [3].

References

- [1] Links stock - <http://sape.ru>.
- [2] Zosimov V., Stepashko V., Bulgakova O. Enhanced technology of efficient Internet retrieval for relevant information using inductive processing of search results. / Artificial Intelligence Methods and Techniques for Business and Engineering Applications / G.Setlak, M.Alexandrov (Eds.). – Rzeszow, Poland; Sofia, Bulgaria: ITHEA, 2012. – 345 p. / – P. 99-112.
- [3] Zosimov V., Bulgakova O. Usage of grid systems for the distribution of the computing process by data in the inductive modeling algorithms. - Intellectual systems for decision making and problems of computation intelligence ISDMCI-2013, Ukraine. – Yevpatoria, 2013.