

Processing information system for highly specialized information in corporate networks

**M O Petrosyan, I V Kovalev, P V Zelenkov, VV Brezitskaya,
G A Prohorovich**

Siberian State Aerospace University named after Academician MF Reshetnev
31 “KrasnoyarskiyRabochiy” prospect, Krasnoyarsk, 660014, Russia.

E-mail: mopetrosyan@gmail.com

Abstract. The new structure for formation system and management system for highly specialized information in corporate systems is offered. The main distinguishing feature of this structure is that it involves the processing of multilingual information in a single user request.

Currently, there is an active development of information technologies. One of the most interesting questions while using information technologies is question about how to collect, process and control information. [1] Computer information retrieval systems hold all the more important place in science and education, especially the Internet, which is an extensive reference tool. Development of the Internet technology ensures that the daily increasing number of information resources made available for public use, increasing the volume of thematic and focused information on the various subject areas.

The expansion of the global network and connectivity to it resulted in a significant increase in the number of its users. The increase of users in recent years, are most pronounced in Russia. Since 2006, the number of Internet users has increased by 10 times. A large part of the Russian-speaking segment of the Internet users in information gathering use existing search services of general purpose. By data for June 2015, most popular information retrieval services are: Yandex, Google, Mail, Rambler. (fig.1)

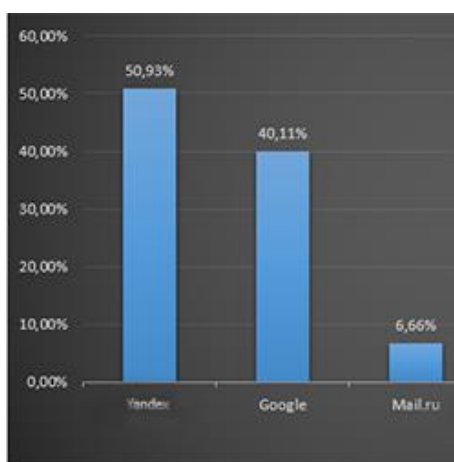


Figure 1. The share of search queries in the Russian-language segment of the Internet

However, it should be noted that these services allow good results only when working with general topics, and there are difficulties in carrying out a highly specialized retrieval.



Moreover, the problem of these systems lay in multilingual presentation of information in the Internet. [1,] Search general-purpose services operate only in the language set on which an a request has been set, but when looking for a highly specialized personalized information you can immediately arrange of multilingual search procedure. [1, 2]

To solve the problems mentioned above could be used existing technologies and approaches, but with emphasis on the processing of multilingual thematic information. In this regard, technology of implementation of information management systems based on multi-agent approach is well proven. Creating a search systems based on agent-based principle allows modification of individual agents, without significant impact on system performance as a whole. Search multi-agent systems are a kind of metasearch systems. metasearch system architecture provides a single point of access to multiple search engines. In other words serves user requests by polling the other user systems, which are completely independent and do not provide any specific information about the contents of their indices or used search methods. This leads to the fact that the construction of metasearch systems have to solve a series of new problems. For example, query languages, often differ greatly and therefore a simplified search language for metasearch system or reformulation of requests for a particular search system is required.

Another problem is the fusion of the responses from the various search systems. Authors propose the following structure of the organization of interaction between agents of multi-agent corporate systems. (fig.2)

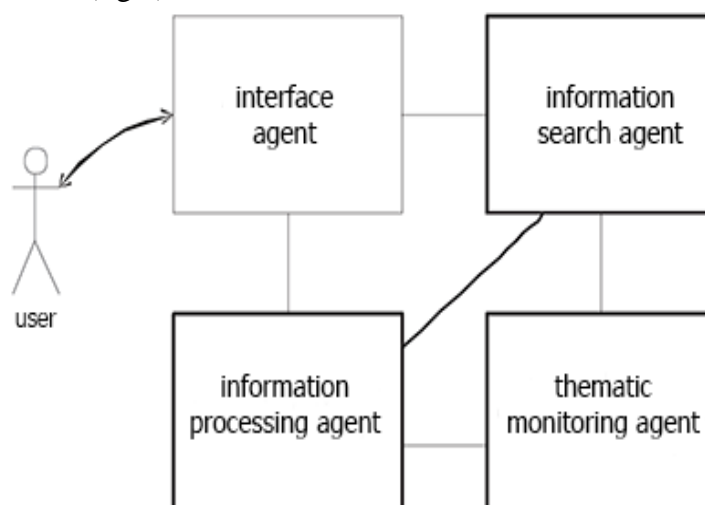


Figure 2. Scheme of multi-agent system

This system consists of four logically related software modules (agents), the purpose and structure of each of which will be shown below. It is connected with two other agents - search agent and processing agent. This agent is simple in structure and performance and performs all operations while interacting with user. It receives a search query, transmits a request to search agents, outputs the search results to a user, and performs a search process of adaptation to the used search systems. Each search agent interacts with a particular search system. (fig.3) This agent requires a more detailed description, as its implementation is proposed to make in metasearch multilingual performance.

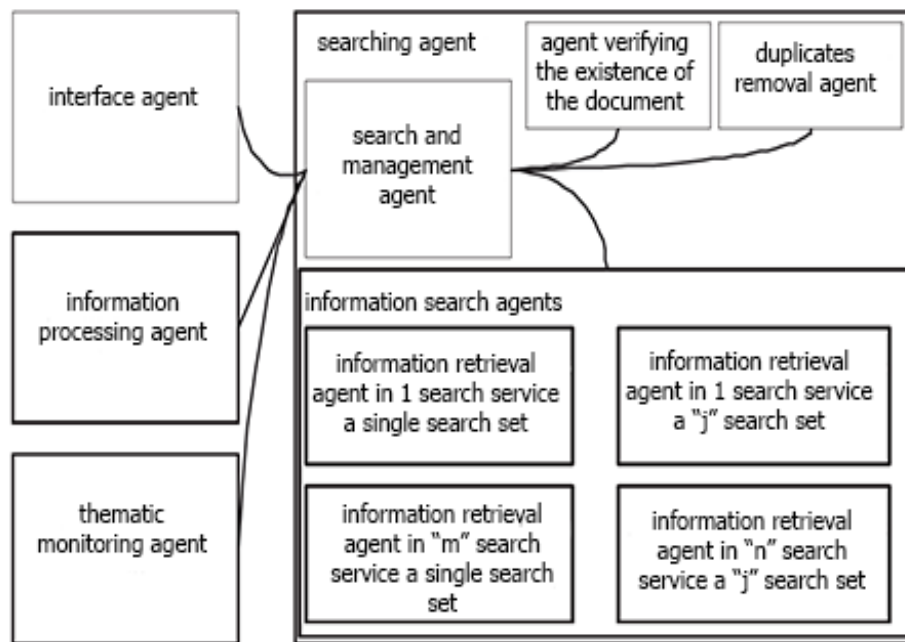


Figure 3. Scheme of the search agent

Search agent is primary. Its main task - is to process user search line that the agent receives from the interface agent. Once the line has been processed, you must initialize the process of multilingual metasearch in the corporate network and the Internet. Next, a process of verification of the existence of documents and removal of duplicate documents is performed. Then, all received data sample is transferred to the information processing agent. (fig.4) This agent is responsible for the information management of the thematic collection obtained in the search stage (from the point of view of the user of corporate system).

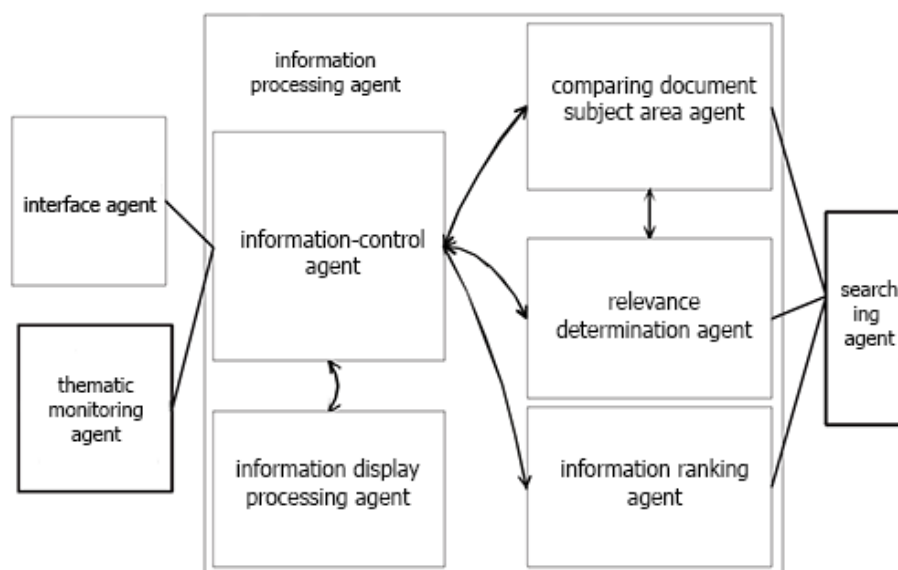


Figure 4. Scheme of information processing agent

Information processing agent consists of:

- information-control agent (this is functionally the main agent of the procedure);
- two agents, which are rigidly connected to each other: relevance determination agent and comparing document subject area agent;
- information ranking agent;
- information display processing agent.

Let us consider in more detail relevance determination agent and comparing document subject area agent. The first agent holds definition of relevance of the documents from the proposed sample. Using algorithms to determine relevance, it can be shown that some documents are more relevant to the request, and some - less. Thus, the problem of processing conditionally relevant documents arise, in other words documents from related subject areas. Therefore, when searching is necessary to determine the possibility of falling into the resulting data sample from such areas. And here we must take into account system user preferences and solve the problem of the inclusion of documents from related subject areas in the result set (or exclusion from it). That is the problem and solves relevance determination agent and comparing document subject area agent. In addition, in the chosen text of the selected subject area relevant can be not the entire document, but only a part, such as separate sections of the general purpose books, selected articles from collections of articles, sections of reports, and organizations and so on. Taking account of this limitation will help in deciding whether to grant the user only the necessary pieces of information.

The next agent - a ranking agent. It is no less important in the processing of information, since when issuing to the user several thousands of documents in the first place on the display list must be the most valuable documents. Thematic monitoring agent is responsible for the analysis of corporate system user information preferences within the information and thematic collections and providing him personalized navigation support and personalized data. We reduce time needed to find the right information, and user traffic in both corporate and in the external network to see only high-quality information by providing users with information collections of personalized navigation menu links to pages that are close to their thematic preferences. Thus, the proposed solutions should enhance the user experience with the corporate system of information resources and serve as an additional incentive for them to more frequent visit these resources. This approach would also greatly reduce the burden on both the inner (corporate) and the external traffic.

References

- [1] Karaseva M V, Bachurina E P, Zelenkov P V, Brezitskaya V V 2015 Information search module based on multilinguistic thesauruses *IOP Conference Series: Materials Science and Engineering* Volume **70** Issue 1 Article number 012011
- [2] Kovalev I V, Polyanskiy K V, Zelenkov P V, Brezitskaya V V, Sidorova GA 2013 System of search, analysis and processing of multilinguistic texts, integrated with data retrieval systems *Vestnik SibGAU* Volume: **1 (47)** P. 48-52.