SYSTEM FOR PROCESSING OF HIGHLY SPECIALIZED INFORMATION IN CORPORATE NETWORKS: EDUCATIONAL PERSPECTIVE

Prof. Dr. Alena Stupina,

Assoc. Prof. Dr. Olga Almabekova,

Dr. Pavel Vcherashny,

Assoc. Prof. Dr. margarita Karaseva,

Assoc. Prof. Dr. Irina Bagdasarian,

Siberian Federal University, Russia

ABSTRACT

Meeting the need to be up-to-date with the latest innovations in the area of specialization modern university students and teachers has been becoming more and more technically savvy. Still, the majority of Internet users employ the existing search services of general purpose to collect information. However, these services provide good results only when working with themes of general public profile. At the same time, the amount of subject-oriented information is constantly growing: it results in arising problems in the effective specialized information search.

The paper offers a new structure of the search system development and control in case of the need to find highly specialized information in corporate systems. The main peculiarity of this structure is that it assumes multilingual information processing within the bounds of one user's inquiry.

Within the multi-agent approach, the agents of the relevance determination and matching the document to the subject area are considered in detail. The proposed solution can be helpful for both students' individual work and teachers' professional development, which assume searching highly specialized information in corporate systems.

Keywords: Multi-agent system, meta-search, highly specialized information, corporate system, information control system.

INTRODUCTION

The most important problem in information technologies' application is collection, processing and control of all types of information[1], [2]. Internet has become the most popular IT means and a broad reference tool. Internet technologies are rapidly developing: every day an increasing number of information resources is provided for public access and the volume of the subject-oriented information for various subject areas has been growing.

For information collecting the majority of Russian Internet users operate with the existing search engines of general use. The most popular are Google, Yandex, Mail, and

Rambler. They account for 95% of the user requests. The distribution of the number of user requests per each service is shown below (Figure 1).
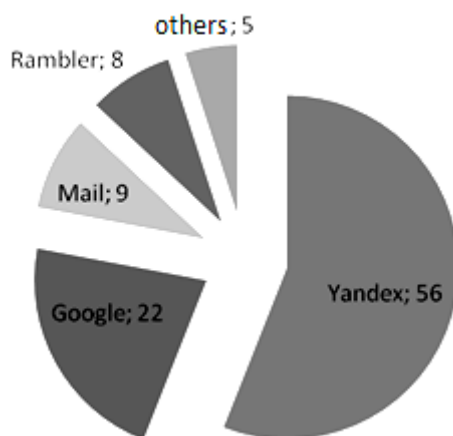

Figure 1 – Share of search queries in Russian segment of the Internet

However, these services provide good results mostly when searching for queries of general themes. However, the challenge arises when the search is subject oriented and specialized information is in demand. Another challenge is the language of information representation[1]. The general search services investigate only the language set on which the search query was specified; however, when searching the highly specialized personalized information one needs multilingual search procedure[1], [2].

METHODS AND MATERIALS

To solve the problems mentioned above the existing technologies and approaches can be applied, but with the emphasis on the processing of multilingual subject-oriented information.

This paper proposes applying the well-proved technology of the information control systems implementation based on the multi-agent approach. The multi-agent search systems are some sort of metasearch engines. The operation is based on the agent principle, which allows modifying individual agents without making a significant impact on the operability of the system as a whole. The architecture of a metasearch engine provides a unified access to multiple search systems, i.e. it serves the user requests by querying other user systems that are completely independent and do not provide any specific information about the contents of their indices or search methods used. This can lead to new problems: for example, the query language used in another search system can often be different, and therefore it requires either a simplified search language for the metasearch engines or the reformulation of the queries for each search system. Another problem is the responses merging from different search systems.

The structure of the interaction between agents of the proposed multi-agent corporative system can be presented as follows (Figure 2).
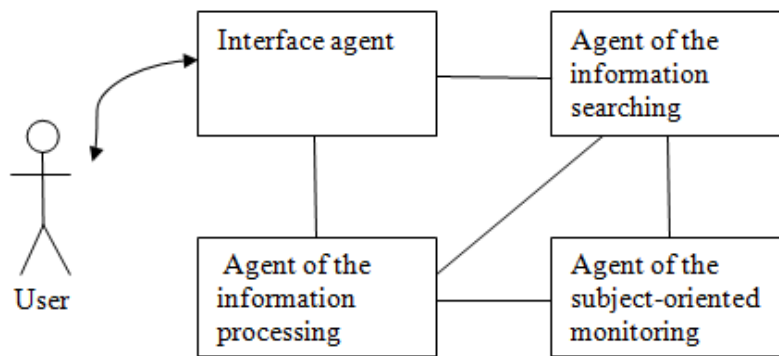
Figure 2 – Generalized scheme of the proposed multi-agent system

As one can see in the structure of the proposed system, it consists of four logically related software modules (agents). The purpose and structure of each of them will be demonstrated below.

The flowchart shows that the interface agent is responsible for the organization of user's information processing system and is associated with two agents (search agent and the processing agent information). The agent is simple in structure and operation. The interface agents perform all operations on user interaction: a search request transmit the search request to the agents, provide output to the user, performing the adaptation of the search process used to search systems.

Each search agent interacts with a specific search system. An agent that searches for the information requires a more detailed description as it is proposed to make its implementation in a multilingual metasearch performance. The structure of this agent is presented in Figure 3.
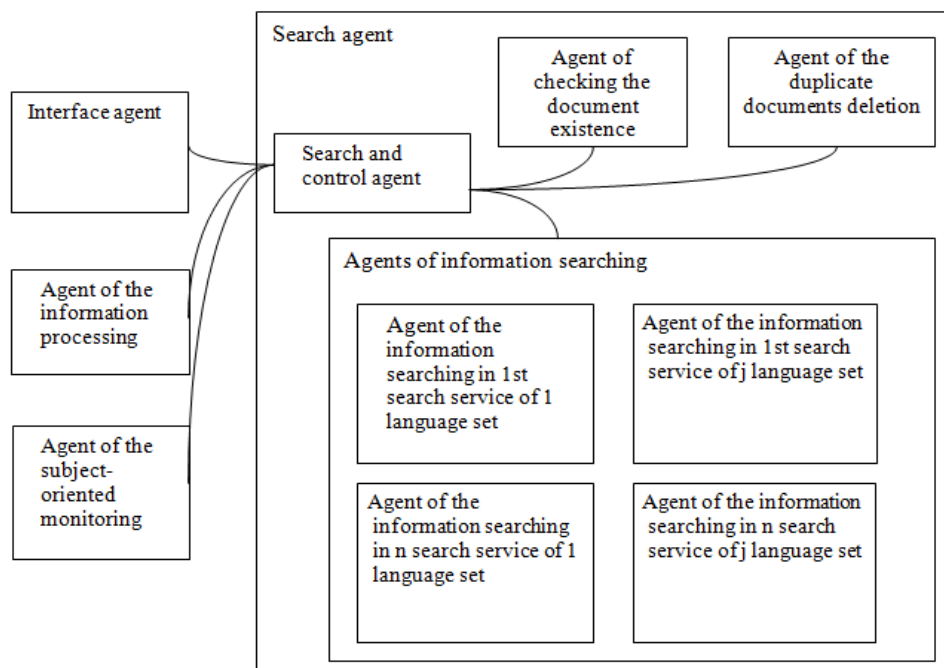

Figure 3 – Diagram of search agent operation

We can consider the structure of the agent as primary. Its main task is processing of the user search line which is received from the interface agent. After the line has been processed multilingual metasearch procedure is initialized both in the corporate network and in the Internet. Then, checking of the documents availability and duplicate documents deletion are performed. After that the produced sample of the information is transferred to the information processing agent. The structure of this agent is given in Figure 4.
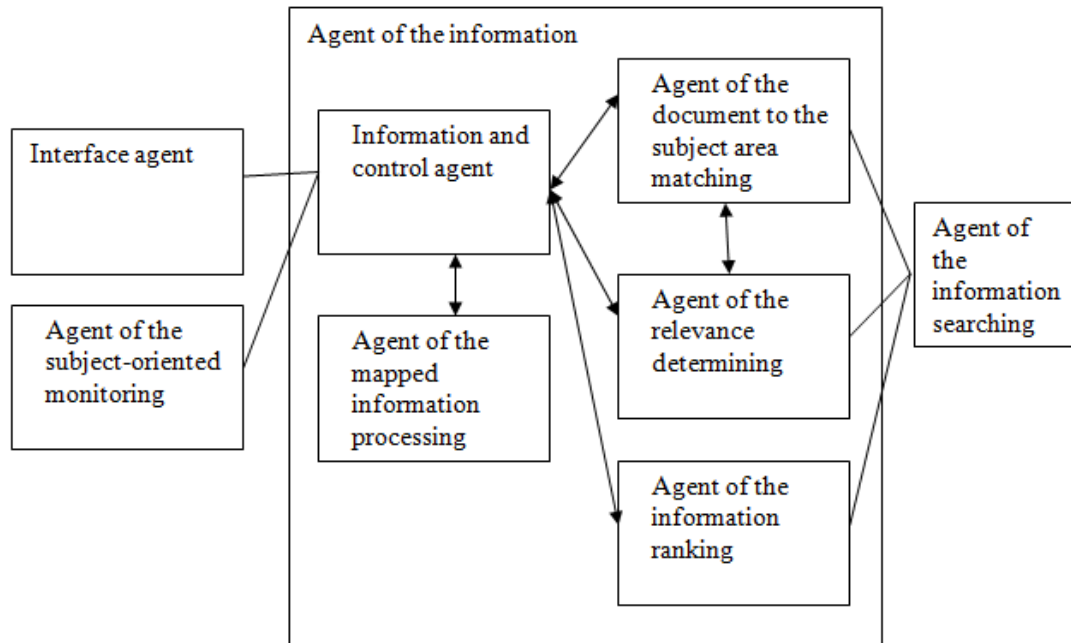

Figure 4 – Diagram of the information processing agent

This agent is responsible for the control of information thematic collections created during the search by the user of the corporate system.

As one can see in the figure, the agent contains the following parts:

- information and control agent (functionally the main agent of this procedure);

- two agents closely connected with one another (an agent of the relevance determination and an agent of the document to subject area matching);

- agent of the information ranking;

- agent of the mapped information processing.

Let us consider in more detail the agents of relevance determination and agents of matching the document to the subject area. The agent of the relevance determination identifies if the proposed sample is relevant to the documents. Having applied the algorithms of the relevance determination we can see that some documents are "more relevant to the query", and some of them are "less less relevant to the query". This arises the problem of processing documents from the related subject areas. Therefore, the search is to consider the possibility of falling in the sample output data of the subject areas. Having known the proximity to data subject areas, user preferences system and solve the problem of inclusion or exclusion of documents from related subject areas in

the sample output are also to be considered. We will show below how this problem can be solved by the agent of the information mapping of the subject area. Another challenge is that not the whole document but only a part of it or separate sections of textbooks of general purpose, separate papers from the collections of papers, chapters from reports of organizations, are relevant to the subject area. This restriction leads to the need to make a decision on supporting a user with only the target necessary part of information.

Another agent to be considered in more detail is a ranking agent. It is important because the most valuable documents are to be mapped first while the information processing since the retrieval of several thousand of documents.

RESULTS

The agent thematic-oriented monitoring is responsible for the analysis of preferences information of the corporate system user as a part of the subject collections and the provision of personalized navigation and personalized data. The time required for the user to find the relevant information, user traffic both in the corporate and external network by viewing only qualitative information are reduced due to the support with the information collections of the personalized navigation menu of links on the page that are close to their thematic preferences.

CONCLUSION

The majority of Internet users employ the existing search services of general purpose to collect any type of information. However, these services provide good results only when working with themes of general public profile. This research was focused on meeting the user need to be able to conduct subject-oriented search.

The paper offers a new structure of the system of highly specialized information search in corporate systems. The main specific feature of this structure is multilingual information processing of users queries under multi-agent approach of relevance determination and matching the document to the subject area.

The proposed solutions can used in students' individual work and teachers' professional development, as well as in corporate computer systems for professional users.

Thus, the proposed solution is aimed at raising the usability of information resources for education and corporate systems and can serve as an additional motivation for students and professional to use these information collections foe educational, research and professional purposes more often.

Most importantly, the time required for the user to find the relevant information, user traffic both in the corporate and external network by viewing only qualitative information are reduced due to the support with the information collections of the personalized navigation menu of links on the page that are close to their thematic preferences.

REFERENCES

[1] Zelenkov P. V. Multilingual model of a distributed system based on thesaurus / P. V. Zelenkov, V. S. Rogov, I. V. Kovalev, M. V. Karaseva // Bulletin of Siberian state

aerospace University: collection of scientific. works. – Krasnoyarsk: Siberian State Aerospace University, 2008. – № 1(18). (in Russian)

[2] Kovalev I.V. Model of the reliability analysis of the distributed computer systems with architecture "client-server" / Kovalev I.V., Zelenkov P.V., Karaseva M.V., Tsarev R.Y. IOP Conference Series: Materials Science and Engineering 17. Vol. "XVII International Scientific Conference", 2015. (in Russian)