

Formation and control system by the specialized data in information networks

I V Kovalev, P V Zelenkov, M V Karaseva, V V Brezitskaya, D I Kovalev
Reshetnev Siberian State Aerospace University
31 “Krasnoyarsky Rabochy” prospect, Krasnoyarsk, 660037, Russia.

E-mail: zelenkow@rambler.ru

Abstract: The paper considers the problem of collection, processing and control by the highly specialized information. The system offers the correct solutions for a multilingual representation of information in the Internet. It is proposed to use the implementation technology of the information control systems based on multi-agent approach.

Nowadays an active development of the network technologies occurs. The most important point in these technologies remains a problem of collecting, processing and information control. The majority of users in the Russian segment of the Internet use the existing retrieval services while the information collecting. According to the information on November, 02 2015 the most popular information search services are as follows: Google, Yandex, Mail, Rambler which account for 95% of the user queries. The distribution of the user queries number can be given in the follow way (Fig. 1).

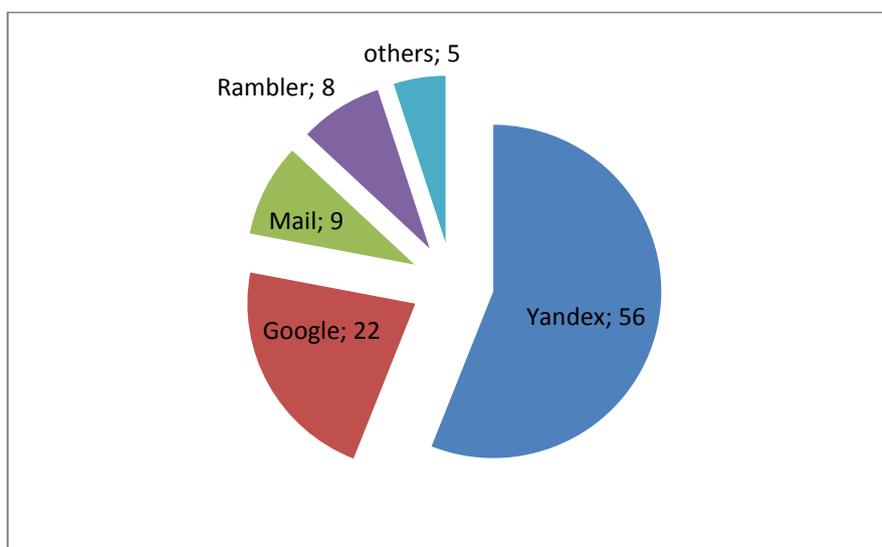


Figure 1. The share of retrieval queries in Russian segment of the Internet.

However, it should be noted that these services give good results while working with general subjects [1]. But if you need to retrieve specialized information you may have difficulties. Also the problem of multilingual representation of information in the Internet can

be correctly solved in these systems. The retrieval services of the general purpose retrieve only in the language set the specified retrieval query was conducted. However, when searching the highly specialized personalized information one can arrange a multilingual retrieval procedure.

To solve the described above problems it is offered to use the existing technologies and approaches but with the emphasis on the processing of subject-oriented information.

First, it is offered to use the proven technology of the information control systems implementation based on the multi-agent approach.

The considered structure of interaction among agents of the offered multi-agent system can be shown as follows (Fig. 2).

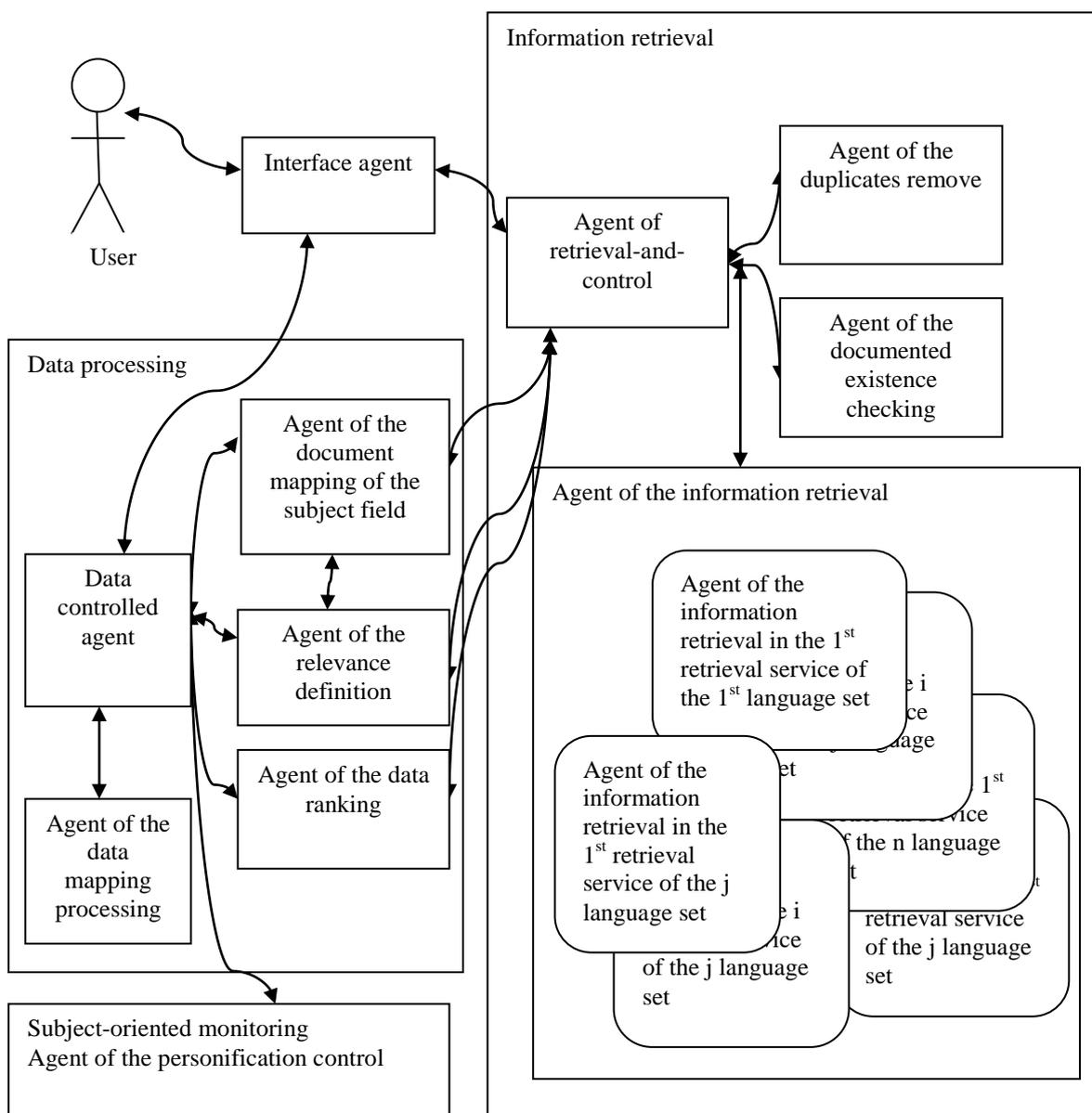


Figure 2. General scheme of the offered multiagent system

As one can see in the structure of the considered system it is necessary to design three logically-related software modules [2]:

1. The first module is responsible for the information quality in the Internet retrieval (formation of multilingual subject-oriented collections). In addition, it is necessary this module to define the relevance of the documents from a thematic catalogs (an agent of the relevance determination). Using the algorithms of the relevance determination it is possible to show that some of the documents are in some way "like a more relevant queries", and some of them are "less relevant". Thus we have a similar complex problem for the user associated with the problem of the documents delivery (an agent of the data ranking). A significant problem that this module solves is the problem of identifying arbitrary relevant documents (documents from the related subject fields). Therefore, a user must take into account the possibility of falling into the resulted sample of the subject areas. If you know the proximity of these fields, you should take into account a user preference of the system and solve the problem of inclusion or exclusion of the documents from the related subject fields into the resulted sample (an agent of the document mapping of the subject field who works closely with the agent of the relevance definition). In addition, only a part of the whole document can be in the selected text of the subject field, for example textbooks of a general purpose, collections of papers, etc. Therefore it is necessary to make a decision to ensure this user with only the necessary information.

2 The second module is responsible for the thematic collection control from the point of view of the user of this system. The main problems here is information processing from adjacent subject fields of the thematic collections (for example, the information on systems analysis and higher mathematics), as well as the storage and processing of the multilingual information in the framework of subject-oriented collections. Also the problem of the system response ranking to a user inquiry is also significant while this subsystem realization [3].

3. The third module is responsible for the subject-oriented monitoring of the user preference of thematic collections and the provision with the personalized navigation support and personalized data based on the identified thematic preferences. Providing users with the information collections of the personalized navigation menu of links into the pages that are close to their thematic preferences, the time required to retrieve the relevant information will; reduce, and the number of pages they view will increase.

The proposed solution should improve the usability of users with information resources and will serve as the additional motivation for them to visit these services more often: it will lead to the formation of a more constant audience of the thematic-oriented collections in the Internet.

References

[1] Zelenkov P. V., Karaseva M. V., Kovalev I. V. Models and methods for optimization of collecting and processing information // Vestnik of SibSAU. – Vol. 2. – 2012. – p. 69-72.

[2] Karaseva, M.V., Bachurina, E.P., Zelenkov, P.V., Brezitskaya, V.V. Information search module based on multilinguistic thesauruses (2015) IOP Conference Series: Materials Science and Engineering, 70 (1), art. no. 012011. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84921447476&partnerID=40&md5=728ef39ad283268034a48b7245b47dcf>

[3] Kovalev I. V., Karaseva M. V., Leskov O. V. Components of information support for multilingual adaptive-learning technology / Scientific technical journal "Control systems and information technologies", Scientific book, 2009, № 1.3 (35). – S. 360 - 363.