



2014-2016 “Humanitarian: creation of the virtual information and resource center for extraction of knowledge from humanitarian texts on the basis of the advanced full text search and functional integration of resources and services in the distributed environment” (Russian Foundation for Humanities - RFH 14-03-12017) Project Annotation

Project Coordinator: ***Sergey H. Lyapin***

One of tendencies of development of modern information society is use of information search for knowledge extraction from texts that is especially important for texts of a humanitarian orientation (philosophy, cultural science, history). In the framework of the project is the adaptation of the universal electronic library has the capabilities of an advanced full-text search and the various forms of presentation of obtained results.

Such technologies, available participants of the project, allow explicate as "horizontal" microcontexts (within the author's paragraph or a cluster of paragraphs), and "vertical" macrocontexts (within the whole document or set of documents), to make preliminary substantial structuring (to build ontologies) subject domain of any set of documents. In the project possibility of the distributed full text search which is functionally uniting resources of several similar electronic libraries is realized also.

On the basis of these approaches and technologies the virtual information and resource center for extraction of knowledge from humanitarian texts and functional integration of resources and services in the distributed environment is created.

At the first stage (2014) the experimental information environment with the above-named opportunities and resource base in 1000 of Russian-language humanitarian texts (subject domains is created: philosophy, history, cultural science) on the basis of humanitarian department of St. Petersburg university for information technologies, mechanics and optics. The methodology of implementation of the inquiries focused on an explication of various semantic contexts and on a tematization of a material is developed and described.

At the following stages (2015-2016) further development of intellectual opportunities of search (semantic mapping of results of inquiry, development and inclusion of thesauruses in full text inquiry, development and realization of cascade and composite inquiries) will be provided. The number of the participants possessing similar technologies (will be also expanded to 7-8 organizations of education, science and culture) and the available resource base (to 10 thousand Russian-speaking full text resources) is increased on which the distributed full text search with open access from the Internet will be realized.

Prospective Results:

Development of the methodological and technological principles of the organization of the virtual information and resource center of a humanitarian orientation with open access from the Internet and the decentralized "cellular" architecture of creation of the distributed environment.

Keywords: knowledge extraction, semantic context explication, full-text search, distributed information systems, functional integration of resources

Contacts

ITMO University

eGovernance Center & Chair of State Information Systems Management



Office 333, Birzhevaya linia, 14, 199034,
Vasilevsky Island, St.Petersburg, Russia
St.Petersburg National Research University of Information technologies,
mechanics and optics (ITMO University)
E-mail: info@egov-center.ru
Tel./fax: +7 812 457 17 96
<http://egov.ifmo.ru/en>