

ПРОТОТИП МНОГОФУНКЦИОНАЛЬНОЙ ПОЛНОТЕКСТОВОЙ БИБЛИОТЕКИ В АРХИТЕКТУРЕ WEB-BROWSER / WEB-SERVER / SQL-SERVER.

Ляпин С.Х., Куковьякин А.В.

НП “Центроконцепт”, 163002, Архангельск, пр. Ломоносова, 2, оф. 207

E-mail: lyapin@crc.pomorsu.ru

PROTOTYPE OF MULTIFUNCTIONAL FULL-TEXT LIBRARY IN THE ARCHITECTURE WEB-BROWSER / WEB-SERVER / SQL-SERVER

Sergey Lyapin, Alexey Kukovyakin

NP “Centroconcept”, 163002, Arkhangelsk, Lomonosov av., 2, of. 207

E-mail: lyapin@crc.pomorsu.ru

Within the framework of the research program “Textaurus” an operational prototype of multifunctional library T-Libra v.4.1. has been created which makes it possible to carry out flexible parametrizable search within a full-text database. The information system is realized in the architecture Web-browser / Web-server / SQL-server. This allows to achieve an optimal combination of universality and efficiency of text processing, on the one hand, and convenience and minimization of expenses for an end user (due to applying of a standard Web-browser as a client application), on the other one.

The following principles underlie the information system: a) *multifunctionality*, b) *intelligence*, c) *multilingual primary texts and full-text searching*, d) *development of digital library (DL) by a user* (“administrative client”), e) *multi-platform working*.

A “library of concepts”, i.e. a block of functional models of semantic (concept-oriented) searching, as well as a subsystem of parametrizable queries to a full-text database, which is closely connected with the “library”, serve as a conceptual basis of multifunctionality and “intelligence” of the DL T-Libra v.4.1. An *author’s paragraph* is a unit of full-text searching in the suggested technology. At that, the “logic” of an educational / scientific topic or a problem can be built in a multilevel flexible structure of a query and the “library of concepts”, replenishable by the developers and experts.

About 10 queries of various level of complexity and conceptuality are realized in the suggested version of the information system: from simple terminological searching (taking into account lexical and grammatical paradigms of Russian) to several kinds of explication of terminological fields and adjustable two-parameter thematic searching (a [set of terms] and a [distance between terms] within the limits of an author’s paragraph are such parameters correspondingly).

В рамках исследовательской программы “Textaurus” создан действующий прототип многофункциональной библиотеки T-Libra v.4.1. в среде Windows с гибким параметризуемым поиском по полнотекстовой базе данных. Информационная система реализована в архитектуре Web-browser / Web-server / SQL-server. Это позволяет достичь оптимального сочетания, с одной стороны, универсальности и эффективности текстового процессинга и, с другой стороны, удобства и минимизации затрат для конечного пользователя (стандартный Web-браузер в качестве клиентского приложения). В основу разработки положены следующие принципы: а) *многофункциональности*, б) *интеллектуальности* (в архитектуре ЭБ предусмотрены: блок электронных словарей с программой автоматической генерации лексико-грамматических парадигм естественного языка, блок функциональных моделей концепт-ориентированного поиска, своеобразная “библиотека концептов”, блок параметризуемых поисковых запросов), в) *многоязычности первичных текстов и полнотекстового поиска*, г) *развития ЭБ силами пользователя* (‘административного клиента’), д) *многоплатформенности*.

Многофункциональность. Технологически ЭБ T-Libra v.4.1. объединяет три функциональные подсистемы: а) подсистему унифицированных полнотекстовых SQL-баз данных, находящихся под управлением эффективной промышленной СУБД; б) настраиваемую подсистему “интеллектуального” (концепт-ориентированного) полнотекстового поиска; в) настраиваемую подсистему представления результатов поиска в виде, адаптированном к характеру текущей деятельности конечного пользователя. Гибкая организация взаимодействия этих подсистем определяет многофункциональность ЭБ, то есть возможность ее использования не только по своему прямому назначению (хранение и презентация полнотекстовых электронных ресурсов), но также в качестве *обучающей* и *исследовательской* среды, – постольку, поскольку в основе обучения и исследования лежит многоплановая работа с ‘текстом’.

“Интеллектуальность” и многоплановая обработка текста. Содержательной основой многофункциональности и интеллектуальности ЭБ T-Libra v.4.1. является “библиотека концептов”, то есть блок функциональных моделей смыслового (концепт-ориентированного) поиска, а также тесно связанная с ним подсистема параметризуемых запросов к ПТБД. Единицей полнотекстового поиска в предлагаемой технологии выступает *авторский абзац*. При этом ‘логика’ учебной / научной темы или проблемы может быть встроена в многоуровневую гибкую структуру запроса и пополняемую разработчиками и экспертами ‘библиотеку концептов’. В предлагаемой версии реализовано около 10 запросов разной степени сложности и концептуальности: от простого терминологического поиска до нескольких видов экспликации терминологических полей и настраиваемого двухпараметрического тематического поиска (параметрами являются соответственно [множество терминов] и [расстояние между терминами] в пре-

делах авторского абзаца). Научной основой разработки является понимание термина как редуцированного смыслового концепта, и, соответственно, восстановление смыслового (микро)контекста по эксплицированным терминологическим полям.