

Поддержка гуманитарных исследований сервисами полнотекстового поиска в распределенной среде (проект «Humanitarianiana») *

С.Х. Ляпин, А.В. Куковякин, И.И. Толстикова

ООО «Константа», Архангельский краеведческий музей, Университет ИТМО
lyapins@yandex.ru, magus@softconst.ru, tolstikova_irina@mail.ru

Аннотация

Обосновывается использование сервисов продвинутого полнотекстового поиска (в том числе в режиме распределенной информационной среды) для поддержки гуманитарных исследований. В этой связи описывается проект «Humanitarianiana» по созданию виртуального информационно-ресурсного центра для извлечения контекстного знания в гуманитарной сфере, создаваемого на базе Университета ИТМО (г. Санкт-Петербург) с партнерами.

Доклад подготовлен при поддержке гранта РГНФ № 14-03-12017.

1. Введение

Важными тенденциями развития современной информационной среды являются: а) создание многофункциональных электронных библиотек, предоставляющих сервисы продвинутого полнотекстового поиска [1], [2] и б) создание распределенных информационных систем, обладающих соответствующим функционалом [3], [4].

Сервисы полнотекстового поиска являются универсальным инструментом поддержки различных исследований, особенно в социогуманитарной сфере. Но для этого они должны быть достаточно «продвинутыми», т.е. а) доставать не только до текста в целом, но и до произвольных фрагментов текста, б) давать возможность экспликации контекстного знания, в) иметь механизмы гибкой тематизации и фокусировки соответствующих запросов.

Распределенная среда с возможностями полнотекстового поиска позволяет существенно увеличить количество доступных ресурсов для целевых групп пользователей, а также создавать на этой технологической базе виртуальные профессиональные сообщества для поддержки коллективных исследований [5]

В докладе описывается и демонстрируется одна из возможных реализаций такого подхода,

технологической основой которой является информационная система T-Libra (разработка ООО «Константа», Архангельск).[6]

В настоящее время библиотеки на основе ИС T-Libra функционируют в Музеях Московского Кремля [7]; в Архангельском краеведческом музее [8]; в ООО «Константа» [9]; в общедоступных библиотеках (ЦГПБ им. В.В.Маяковского, Санкт-Петербург; библиотека «Дом А.Ф.Лосева» (Москва); библиотека киноискусства им. С.М.Эйзенштейна (Москва); научно-образовательных организациях России (Университет ИТМО, Санкт-Петербург), участвуют в различных проектах по созданию современной сервис-ориентированной информационной среды.

В п.2 рассматриваются основные сервисы полнотекстового поиска в ИС T-Libra, приведены примеры экспликации контекстного знания для нескольких тем гуманитарной направленности. В п. 3 вышесказанное рассматривается применительно к *распределенному полнотекстовому поиску* и созданию виртуального информационно-ресурсного центра для извлечения контекстного знания в гуманитарной сфере, создаваемого на базе Университета ИТМО (Санкт-Петербург) с партнерами (проект «Humanitarianiana»). Обозначены также ближайшие перспективы развития предлагаемой технологии.

2. Сервисы полнотекстового поиска для поддержки гуманитарных исследований

2.1. Архитектура и основной функционал ИС T-Libra

Информационная система T-Libra v.6.x. предназначена для создания многофункциональных многоязычных электронных полнотекстовых библиотек с возможностями гибкого тематизируемого полнотекстового поиска.

ИС T-Libra разработана в клиент-серверной Интернет/Интранет архитектуре. Для функционирования на сервере требуется операционная система семейства Windows: 2003/2008/XP/Vista/7, на клиентском месте – Windows / Linux. На сервере также используется

свободно-распространяемое ПО: в качестве веб-сервера – Apache, в качестве СУБД – реляционная система управления базами данных MySQL.

Клиентской программой для T-Libra является веб-браузер (рекомендуются Mozilla Firefox, Opera, Apple Safari, Google Chrome, Яндекс.Браузер). В этом смысле T-Libra 6.x является многоплатформенной. ИС T-Libra является многоплатформенной также в плане совместимости ее серверной части с различными СУБД, поддерживающими протокол ODBC (Open Database Connectivity).

Имеется поддержка распределенного полнотекстового поиска в однородной информационной среде, включающей в себя несколько ИС T-Libra, взаимодействующих между собой по Интернет-протоколам.

Информационная система T-Libra позволяет реализовать следующие типы полнотекстового поиска:

Абзацно-ориентированный: в документах, включенных пользователем в поисковую область («корзина ресурсов»), находится множество абзацев, удовлетворяющих условиям запроса (тем самым эксплицируется «горизонтальный» микроконтекст, в котором в составе абзаца находятся искомые термины). Авторский абзац выбран в качестве естественной единицы смыслового членения текста.

Частотно-ориентированный: создает частотно-ранжированный список терминов (имен существительных) из документов на заданную глубину с указанием абсолютной и относительной (в промилле) частоты встречаемости термина. В получаемой «терминограмме» тем самым эксплицируется «вертикальный» макроконтекст, характеризующий предметную область документа в целом или их выбранной совокупности.

Каждый из этих типов поиска может включать в себя несколько разновидностей.

В рамках распределенной информационной среды в настоящее время реализованы **абзацно-ориентированные запросы**, наиболее востребованные для тематической обработки документов. В их числе:

А) Простой («однослойный») тематический поиск, с одним комплексным полем для ввода терминов и использованием для этих терминов операторов логического *объединения*, обязательного *исключения* или обязательного *включения* термина в запрос. Результатом поиска является список абзацев, удовлетворяющих заданным условиям.

Каждый из абзацев, входящих в результаты запроса, может быть одним «кликом» мышки раскрыт до своего полного вида. Используя опцию «**Контекст**» в левом меню, можно последовательно раскрыть абзацы до и после найденного – вплоть до кластера из семи абзацев (три абзаца «до», три абзаца «после», плюс сам абзац – результат запроса).

Б) Расширенный («многослойный») тематический поиск. Этот вид поиска сохраняет весь функционал простого тематического поиска и обладает дополнительными возможностями тематической фокусировки запроса. Соответствующий инструментарий включает в себя: а) формирование нескольких поисковых полей («слоев») и б) включение в запрос дополнительных количественных параметров его фокусировки.

Поисковое поле «слой» представляет собой технический инструмент для выделения того или иного содержательного «аспекта» интересующей пользователя «темы»; всего может быть сформировано от 2 до 8 слоев. Между слоями действует операция логического пересечения (оператор «AND»), внутри слоя – операция логического объединения (оператор «OR») заданных терминов. Имеется возможность комбинировать актуально используемые слои.

Еще более точная тематическая фокусировка запроса достигается за счет выполнения дополнительных условий: а) указания минимально необходимого количества поисковых слоев (от 2 до 8); б) указания максимального расстояния между терминами, принадлежащими разным слоям: от 0, когда слова из двух разных слоев запроса в составе абзаца примыкают друг к другу, до произвольной величины.

Можно обеспечить и более жесткую фокусировку запроса, задав, например, в трехслойном запросе расстояние между терминами = 1 (в четырехслойном = 2 и т.д.), т.е. при таких условиях все поисковые термины в абзаце должны быть расположены рядом друг с другом.

2.2. Примеры экспликации контекстного знания с помощью сервисов полнотекстового поиска

Развивая идею о сервисах информационно-библиотечного полнотекстового поиска для поддержки гуманитарных исследований, на протяжении нескольких последних лет некоторыми из авторов доклада в рамках различных мероприятий и мини-проектов были осуществлены предварительные (пробные) тематические исследования, в частности, по таким крупным темам, как «Проблема факта в русской философии», «Проблема одиночества в русской философии и литературе», «Проблема человека в русской философии и культуре». Эти и аналогичные исследования будут существенно расширены – как по привлекаемой ресурсной базе, так и по изучаемым темам, их аспектам и используемым сервисам – в рамках проекта «Humanitarian».

Для экспликации вышеназванных концептов применялась методика сочетания частотно-ориентированных запросов с абзацно-ориентированными. Для этого вначале строились соответствующие частотно-ранжированные списки терминов (существительных) – «терминограммы» – относительно исследуемых терминов

(соответственно: «факт», одиночество», «человек»). При этом на заданную глубину (как правило, в 10 слов) средствами *относительного частотного поиска* автоматически создавались таблицы частотной встречаемости терминов по выбранной совокупности ресурсов. В итоге мы получали экспликацию предметной области, коррелирующей с исследуемыми терминами/концептами (соответственно: «факт», одиночество», «человек») – т.е. экспликацию «вертикального контекста». Затем термины из этих таблиц включались в абзацно-ориентированные запросы, что позволяло эксплицировать микроконтекст («горизонтальный контекст») употребления этих терминов в рамках авторского абзаца.

Экспликация концепта фактуальности в философии Н.А. Бердяева.

С помощью *частотного поиска* (относительно термина «факт») можно предварительно выявить предметную область (смысловой макроконтекст), в котором существует концепт «факт» в философских текстах Н.А.Бердяева в целом или в тексте отдельного произведения [10]. В частности, для исследуемых работ Бердяева, содержащих термин «факт» (24 произведения) в смысловую окрестность «факта» попадают (т.е. встречаются в абзацах, где обязательно имеется и термин «факт»), в порядке убывания частотности, следующие термины: 1. «жизнь» (850 раз), 2. «мир» (811 раз), 3. «человек» (788 раз), 4. «дух» (484 раза), 5. «Бог» (483 раза), 6. «сознание» (401 раз), 7. «свобода» (377 раз), **8. «факт»** (366 раз), 9. «философия» (324 раза), 10. «история» (311 раз).

Опираясь на эти результаты, эксплицирующие предметную область фактуальности в работах Н.А. Бердяева, можно далее с помощью абзацно-ориентированного поиска осуществить экспликацию горизонтальных микроконтекстов (в пределах авторского абзаца), коррелирующих с концептом «факт».

Можно использовать многослойный (например, трехслойный) запрос. В один слой ввести термин «факт», в два других – термины из бердяевской терминологии (например, «мир» и «человек»), выбрать дополнительные параметры фокусировки запроса (все 3 слоя актуальны, расстояние между терминами не более 8 слов) и получить результат: 7 абзацев в 4-х произведениях (поиск произведен по массиву из 64 текстов Бердяева).

Экспертная оценка этого и других найденных абзацев говорит о высокой степени релевантности результатов таким образом организованного полнотекстового запроса. Эксплицируются микроконтексты концепта фактуальности, связанные с такими аспектами темы как «гуманизм и факт рабства человека у природного мира»; «факт самосознания человека и трансцендентность сознания» и т.д.

Выбирая другие термины из терминологии Бердяева (например, «факт» + «сознание» +

«свобода»; или «факт» + «философия» + «история» и т.д.), включая их в абзацно-ориентированные запросы и варьируя условия их фокусировки, можно эксплицировать самые разные семантические микроконтексты в рамках разработки темы «Проблема факта в философии Бердяева».

С соответствующими изменениями сказанное относится к другим авторам – например, терминологии Лосева и Канта и соответствующим абзацно-ориентированным запросам по их произведениям. Все это дает, в свою очередь, текстологический материал для гносеологической компаративистики.

Аналогичное исследование сделано нами по теме «**Экспликация концепта одиночества в философии Н.А. Бердяева**». Впервые доклад состоялся на международной конференции «Гигиена культуры: актуальные вопросы. Феномен одиночества» (СПбГУ, 20-22 июня 2013 г.). Проведено сравнительное исследование концепта «одиночество» в текстах Н.А. Бердяева, М.К. Мамардашвили, С.Л. Франка. Участниками конференции был отмечен нетривиальный характер получаемых при таком подходе результатов.

С помощью относительного частотного поиска (опорный термин «одиночество») можно предварительно выявить смысловую («вертикальный») макроконтекст, в котором существует концепт «одиночество» в текстах Н.А.Бердяева в целом или в тексте отдельного произведения. В частности, для исследуемых работ (20 произведений, где есть термин «одиночество») в смысловую окрестность этого термина попадают, в порядке убывания частотности, следующие термины (взята первая десятка терминов): 1. «мир» (470 раз), 2. «человек» (384 раза), 3. «жизнь» (334 раза), **4. «одиночество» (308 раз)**, 5. «личность» (296 раз), 6. «общение» (187 раз), 7. «люди» (175 раз), 8. «познание» (170 раз), 9. «общество» (169 раз), 10. «существование» (156 раз).

Опираясь на эти результаты, можно далее с помощью абзацно-ориентированного поиска осуществить экспликацию горизонтальных микроконтекстов (в пределах авторского абзаца), коррелирующих с концептом «одиночество».

Можно использовать трехслойный запрос. В первый слой ввести термин «одиночество», в два других – термины из бердяевской терминологии (например, «мир» и «человек»), выбрать дополнительные параметры фокусировки запроса (все 3 слоя актуальны, расстояние между терминами не более 8 слов) и получить результат: 5 абзацев в 3-х произведениях.

Экспертная оценка этого и других найденных абзацев говорит о достаточно высокой степени релевантности результатов построенного таким образом полнотекстового запроса. В частности, в этом запросе эксплицируются микроконтексты («горизонтальные» контексты) концепта «одиночество», связанные с концептами / темами:

- поиск выхода из одиночества в субъективном мире;
- смерть как абсолютное одиночество;
- символизм глубокого духовного одиночества современного человека;
- замыкание субъективного мира от божественного бытия и обречение тем самым человека на одиночество его душевных переживаний;
- тоска как переживание человеком покинутости, одиночества и чуждости мира.

Выбирая другие термины из терминограмм Бердяева (например, «одиночество» + «жизнь + «личность»), и включая их в соответствующий абзацно-ориентированный запрос, получим следующие эксплицированные микроконтексты:

- через любовь преодолевается одиночество и осуществляется общение;
- личность призвана к творчеству в жизни социальной и космической - духовно личность не одинока и предполагает другого и других, предполагает "ты" и "мы", но никогда не превращается в часть и средство;
- одиночество возникает для "я", когда возникает объект;
- социальная правда общения есть правда о личности, о ее выходе из одиночества;
- человек имеет священное право на одиночество и на охранение своей интимной жизни.
- ошибочно было бы думать, что одиночество есть солипсизм;
- романтизм может принимать формы и крайнего оптимизма, веры в безгрешность человеческой природы и слияния с жизнью космоса, и крайнего пессимизма, одиночества "я", несчастья и трагизма человеческой участи;
- острая тоска возможна в самые счастливые минуты одиночества.

Аналогично для кластера поисковых терминов «одиночество» + «общение» + «познание» получим, например, такие микроконтексты:

- объективация в познании закрывает тайну существования, в которой только и может быть преодолено одиночество и установлено общение.
- С проблемой одиночества связаны проблемы "я", личности, общества, общения, познания.

Включая другие различные сочетания терминов из бердяевской терминограммы в многослойные абзацно-ориентированные запросы и варьируя условия их фокусировки, можно эксплицировать самые разные семантические микроконтексты в рамках разработки темы «Проблема одиночества в философии Бердяева».

Всего в рамках вышеобозначенной модели экспликации – комбинирование частотно-ранжированного и абзацно-ориентированных запросов – мы выделили для последующего исследования около 50 смысловых контекстов употребления термина «одиночество» (в 20 произведениях Н.А.Бердяева).

Mutatis mutandis сказанное относится к другим авторам – например, терминограммам М.К. Мамардашвили и С.Л. Франка и соответствующим абзацно-ориентированным запросам по их произведениям.

Методология, технология и результаты этого исследования были представлены также на одном из заседаний Круглого стола конференции Телематика-2013 (Университет ИТМО, 26 июня 2013 г.)

В духе вышеназванных методологических подходов и технологий начата разработка крупной комплексной междисциплинарной темы «*Проблема человека в русской философии и культуре*» (в рамках тематики исследований кафедры культурологи и межкультурных коммуникаций Университета ИТМО).

Все такого рода исследования предполагают, разумеется, существенное увеличение ресурсной базы (от примерно 2 тысяч документов в настоящее время до нескольких десятков тысяч в рамках развития проекта «Humanitarian» – за счет роста ресурсной базы каждого из его участников и масштабирования проекта в рамках децентрализованной распределенной среды).

3. Полнотекстовый поиск в децентрализованной распределенной среде, управляемой пользовательским браузером

При выборе модели организации распределенной среды для сервисов полнотекстового поиска мы ориентировались на Веб-сервисы и Интернет-протоколы.

Это связано как с архитектурой информационной системы T-Libra (клиент-серверная архитектура, в которой в качестве клиента выступает Веб-браузер), так и с очевидной необходимостью учитывать тенденции развития современного информационного пространства.

В этой связи была выбрана модель децентрализованной среды под управлением пользовательского браузера, обращающегося к множеству независимых серверов, находящихся в ведении различных организаций. Этот подход вполне укладывается в парадигму распределенных информационных систем [11], достаточен для наших целей и позволяет не рассматривать более сложные варианты построения распределенных систем, связанные, например, со взаимодействием унифицированного протокола Z39.50 [12], [13] с Интернет-протоколами [14].

Наш подход может быть рассмотрен как вариант metasearch engine [15]. Инструментом машины метапоиска является в этом случае пользовательский браузер. Взаимодействия серверов участвующих в распределенном поиске организаций при этом не требуется.

3.1. Эксперимент по реализации распределенного полнотекстового поиска в однородной информационной среде

5 апреля 2011 года, в рамках ежегодной конференции «Музейные библиотеки в современном мире», проходившей в Музеях Московского Кремля, был осуществлен эксперимент по реализации распределенного полнотекстового поиска. В нем участвовали электронные библиотеки 5 организаций: 2 в Архангельске и 3 – в Москве. В каждой из них была установлена ИС T-Libra.

Вход в распределенную среду осуществлялся с любого из серверов-участников (фактически были использованы 3 из 5 возможностей). Тайм-аут для отклика серверов был установлен в 10 секунд (параметр регулируется пользователем при формировании запроса), этого оказалось достаточно для успешного ответа всех библиотек. Были осуществлены оба варианта абзацно-ориентированного поиска (однослойный и трехслойный полнотекстовые запросы), продемонстрирована связь найденных абзацев с релевантными графическими страницами соответствующего документа, а также реализована опция Собрать тему (оценивались найденные абзацы и собирались темы по запросам «Собор Василия Блаженного» и «Иконостас Успенского собора»).

Этот эксперимент был воспроизведен во время видеоконференции, состоявшейся через несколько дней в Российской ассоциации электронных библиотек [16], а в 2012-2014 гг. продемонстрирован (с различным составом участников) еще на нескольких крупных конференциях (Президентская библиотека, октябрь 2012; Крым, июнь 2013; Телематика, июнь 2013; Телематика, июнь 2014).

3.2. Алгоритм обработки распределенного полнотекстового запроса

Предварительное замечание. «Базовый» сервер - это сервер, который сформировал страницу, на которой пользователь нажал кнопку "Создать распределенный запрос". Базовым он является функционально, то есть для этого конкретного запроса. Практически это тот сервер, с которым пользователь начал работу с T-Libra в одной из библиотек, включенных в распределенную среду.

1) клиентская часть T-Libra (функционирующая в браузере пользователя), работающая с каким-либо сервером T-Libra («базовый» сервер), получает от базового сервера список адресов других серверов, которые будут участвовать в полнотекстовом запросе (этот список ведётся администратором базового сервера T-Libra).

2) после заполнения формы полнотекстового запроса пользователь инициирует выполнение запроса (кнопкой «Выполнить запрос»).

3) браузер рассылает http-запросы (на выполнение полнотекстового запроса) на базовый сервер и прочие, перечисленные в списке.

4) каждый сервер выполняет запрос, формирует результат и отправляет в клиентскую часть 10 лучших ответов и общий размер результата.

5) из полученных частичных «топ 10» (от каждого сервера) клиентская часть гарантированно строит и отображает пользователю «топ 10» ответа в целом (первая страница результата) и суммарный размер результата.

6) в ожидании реакции пользователя клиентская часть рассылает http-запросы (на пересылку остатков результата) на все серверы, участвующие в выполнении запроса.

7) клиентская часть принимает ответы от них, объединяя частичные результаты и позволяя пользователю просмотреть следующих страниц результата распределенного полнотекстового запроса.

Этот уже реализованный алгоритм будет совершенствоваться по мере развития и масштабирования проекта.

3.3. Проект «Humanitarian» (краткое описание)

На основании разработанной ООО «Константа» технологии создания многофункциональных электронных библиотек с сервисами продвинутого полнотекстового поиска (реализованной в информационной системе T-Libra), опыта ее внедрения и эксплуатации в различных организациях России, осуществленных пилотных проектов по использованию поиска для экспликации различных смысловых контекстов и успешной реализации экспериментов по созданию децентрализованной распределенной среды под управлением пользовательского браузера в 2013 году был инициирован проект «Humanitarian» от имени ООО «Константа» и Санкт-Петербургского научно-исследовательского университета информационных технологий, механики и оптики (Университета ИТМО).

Этот проект получил поддержку РГНФ (грант № 14-03-12017). Он рассчитан на три года (2014-2016 гг.), и осуществляется поэтапно: сначала на организационно-технологической базе его инициаторов (ООО «Константа» и Университет ИТМО), а затем – за счет привлечения партнеров из числа общедоступных и специализированных библиотек, музейных, архивных, образовательных и научных организаций.

Децентрализованная архитектура распределенной среды, не требующая непосредственного взаимодействия серверов ее участников (поскольку управление поисковыми сервисами осуществляется пользовательским браузером), учитывает интересы каждой из организаций-участников (ресурсы остаются на ее серверах и в ее ведении) и обеспечивает ее масштабирование и гибкую настройку.

В рамках проекта планируется создание виртуального информационно-ресурсного центра для извлечения знаний из гуманитарных текстов на основе продвинутого полнотекстового поиска и функциональной интеграции ресурсов и сервисов в распределенной среде.

При координирующей роли Университета ИТМО (г. Санкт-Петербург) поэтапно создается междисциплинарная информационная распределенная среда с открытым доступом.

В рамках проекта разрабатывается типология задач автоматизированного извлечения контекстного знания из гуманитарных текстов, создаются методики составления запросов разного типа и вида для типовых задач извлечения знаний, некоторые из них будут реализованы в технологиях запроса в ходе реализации проекта.

Предполагаемое количество доступных полнотекстовых ресурсов для реализации задач проекта к концу 2016 года – не менее 10000 полнотекстовых документов (книг, статей, трудов конференций, аналитических обзоров, научно-методических разработок по тематике проекта и т.п.).

3.4. Кластеры полнотекстового поиска

Проект «Humanitarian» является одним из примеров структурирования информационного пространства на основе методологии децентрализованных распределенных библиотек, полнотекстовый поиск в которых происходит под управлением пользовательского браузера.

Другими известными нам текущими проектами такого рода являются:

А) создание межмузейной электронной библиотеки с распределенным полнотекстовым поиском. Этот проект стартовал 24-25 октября 2012 года на региональном семинаре «Музей в современном информационном пространстве» в Архангельске. Его инициаторы и непосредственные участники: Музеи Московского Кремля (научно-справочная библиотека), Архангельский краеведческий музей (научная библиотека), ООО «Константа». Часть этого проекта вошла в областную целевую программу сферы культуры Архангельской области на 2013-2015 гг.; в ее рамках предусмотрено создание межмузейной распределенной библиотеки Архангельской области с участием библиотек не менее 10 областных и муниципальных музеев.

Функционал каждой из библиотек и распределенной среды в целом будет развиваться в направлении многофункциональности и мультимодальности, что необходимо для поддержки основной деятельности музеев: информационного сопровождения экспозиций и выставок (отбор и подготовка материала для музейных этикеток и аналитических описаний экспонатов), подготовки и проведения экскурсий, обеспечения научно-методической и научно-исследовательской работы в музее.

Б) Под руководством ЦГПБ им. В.В.Маяковского (Санкт-Петербург) создается распределенная полнотекстовая корпоративная сеть общедоступных библиотек (КСОБ) Санкт-Петербурга. В нее поэтапно будут включены районные централизованные библиотечные системы. Первый рабочий фрагмент распределенной КСОБ (3 участника) планируется к пуску в эксплуатацию осенью 2014 года, в рамках мероприятий традиционной конференции «Электронные ресурсы библиотек, архивов, музеев». Ресурсно-сервисная ориентация этой среды – поддержка краеведческой работы библиотек, прежде всего в рамках Петербурговедения, предоставление соответствующих сервисов специалистам, научно-образовательным сообществам и гражданам Санкт-Петербурга и России.

Предлагаемая организация сервисов полнотекстового поиска и децентрализованной распределенной среды может быть использована, как в описанных выше проектах, для создания кластеров информационно-библиотечного поиска междисциплинарного и межведомственного характера, тем самым – для масштабирования технологии, качественного развития информационного пространства, увеличения количества и повышения качества доступных цифровых ресурсов и предоставляемых поисковых и презентационных сервисов.

При этом предполагается (в проекте «Humanitarian» в том числе) развитие функционала электронных библиотек и распределенной среды в целом в направлении интеллектуализации поиска (семантический анализ текста, разработка и включение тезаурусов в полнотекстовый запрос, разработка и реализация каскадных и гибридных запросов, использование технологии программных агентов и т.д.) и мультимодальности – в плане предоставления нетекстовых ресурсов и соответствующих поисковых сервисов (включение, в дополнение к полнотекстовым, аудио- и видеоматериалов, осуществление голосового поиска, обработка потокового видео).

Литература

- [1] Ляпин С.Х. Сервисы электронной полнотекстовой библиотеки для образования, науки и культуры. // Научная периодика: проблемы и решения. 2013. № 2(14). URL: <http://www.dilib.ru/journal/articles/50865.php>; Публикация в eLibrary [Электронный ресурс] <http://elibrary.ru/item.asp?id=19013565/>.
- [2] Ляпин С.Х., Куковякин А.В. Электронная полнотекстовая библиотека для поддержки аналитической и исследовательской деятельности // Интернет и современное общество: сборник научных статей. Труды XVI Всероссийской объединенной конференции «Интернет и современное общество». Санкт-

- Петербург, (IMS-2013), 9 – 11 октября 2013 г., СПб.: НИУ ИТМО, 2013. – С. 163-170. URL: <http://ojs.ifmo.ru/index.php/IMS/article/view/28/29>
 Публикация в eLibrary [Электронный ресурс] <http://elibrary.ru/item.asp?id=21718250>
- [3] Ляпин С.Х. Как пройти в распределенную библиотеку? // Современная наука: актуальные проблемы теории и практики. Серия «Гуманитарные науки». 2012. № 7-8. С.17-21. URL: <http://www.vipstd.ru/nauteh/index.php/--gn12-07/595>; Публикация в eLibrary <http://elibrary.ru/item.asp?id=18279699>.
- [4] Андрианова А.А., Груздев И.А., Куковьякин А.В., Ляпин С.Х. Распределенная ЭБС Российской ассоциации электронных библиотек // Новые информационные технологии и менеджмент качества (NIT&QM'2013). Материалы междунаучной конференции / Редкол.: А.Н. Тихонов (пред.) и др.; ФГАУ ГНИИ ИТТ «Информика». – М.: ООО «АРТ-ФЛЭШ», 2013. – С. 118-123.
- [5] Шварцман М.Е. Электронная библиотека как основа виртуальной исследовательской инфраструктуры. URL: <http://www.aselibrary.ru/blogs/archives/1209/>
- [6] См.: <http://www.softconst.ru/tlibra/>
- [7] <http://www.kreml.ru> → наука → библиотека → T-Libra
- [8] <http://gd.kraeved29.ru/tlibra>
- [9] <http://demo.tlibra.ru>
- [10] Ляпин С.Х. Тексты, контексты, концепты: использование современных систем полнотекстового поиска для философских исследований (на материале русской философии). URL: <http://www.losev-library.ru/index.php?pid=6872>;
- [11] Таненбаум Э., Стеен М. Распределенные системы. Принципы и парадигмы. – СПб.: Питер, 2003. С. 23.
- [12] ISO 23950:1998 Information and documentation -- Information retrieval (Z39.50) - Application service definition and protocol specification. URL: http://www.iso.org/iso/catalogue_detail?csnumber=27446
- [13] Жижимов О.Л., Мазов Н.А.. Модель распределенной информационной системы Сибирского отделения РАН на базе протокола Z39.50 // Электронные библиотеки. - 1999. – Т.2. – Вып. 2. URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part2/zhizhimov>
- [14] Максимов Н.В., Сысойкина М.А. О реализации электронной библиотеки с использованием протоколов HTTP и Z39.50 // Электронные библиотеки. 2002. Т. 5. Вып. 1. [Эл. ресурс] <http://www.elbib.ru/content/journal/2002/200201/MS/MS.ru.html>
- [15] Metasearch engine // http://en.wikipedia.org/wiki/Metasearch_engine
- [16] Презентация проекта «Распределенная межмузейная электронная библиотека».

Видеоконференция Российской ассоциации электронных библиотек (НП ЭЛБИ), 08 апреля 2011 г. URL: <http://www.aselibrary.ru/conference/conference43/conference432039>

Support for Humanities Research by Means of Full-text Search in a Distributed Environment (project «Humanitarian»)

S. Lyapin, A. Kukovyakin, I. Tolstikova

Justifies the use of advanced full-text search (including mode of the distributed information environment) to support humanitarian studies. In this connection describes the project «Humanitarian» to create a virtual information resource center to retrieve contextual knowledge in the Humanities, created on the basis of the University ITMO (St. Petersburg) with partners.

* Работа выполнена при поддержке Российского гуманитарного научного фонда, грант № 14-03-12017.