

Электронная полнотекстовая библиотека для поддержки аналитической и исследовательской деятельности

С.Х. Ляпин, А.В. Куковякин

Архангельский краеведческий музей, ООО «Константа»

lyapins@yandex.ru, magicmagus@yandex.ru

Аннотация

Рассматривается использование сервисов многофункциональной электронной полнотекстовой библиотеки, созданной на основе информационной системы T-Libra 6.6.x (ООО «Константа», Архангельск), для поддержки аналитической и исследовательской деятельности. Приводятся примеры такого использования для поддержки управленческой деятельности (в том числе в рамках программы «Электронное правительство»); информационно-аналитической поддержки деятельности организаций сферы культуры (музей); гуманитарных научных исследований (экспликация смысловых контекстов русской философии).

1. Введение

Электронная библиотека с возможностями гибкого тематизируемого полнотекстового поиска представляет собой *универсальный инструмент* для информационной поддержки различных видов аналитической и исследовательской деятельности.

Разумеется, специализированные системы — например, табличные процессоры типа MS Excel или LibreOffice Calc; компьютерные программы, создаваемые в рамках «обработки естественного языка» (Natural Language Processing) [7], семантический (интеллектуальный) анализ текстов [2] и т.п. — обладают дополнительными возможностями, обычно недоступными для универсальной полнотекстовой библиотеки. Но они зато существенно уступают последней как по объему и содержательному разнообразию хранимой информации, так и по эффективности работы с основным носителем значимой информации — текстом, и к тому же не включают в свой функционал полноценное использование библиографических описаний, сделанных в соответствии с международными стандартами. Отметим, что мы ведем при этом речь о библиотеках, содержащих достаточно продвинутые полнотекстовые сервисы.

К последним относится, например, поиск, позволяющий сформировать *тематическую подборку материала из разных документов*, причем с точностью до произвольных единиц полнотекстовой информации — предложений, авторских абзацев и их кластеров, частотно-ранжированных списков терминов и т.д. Сами эти единицы информации, *тематически связываемые запросом*, могут находиться в разных «документах» информационной системы, а в случае распределенной библиотеки — и на разных пространственно удаленных серверах.

Разумеется, электронные библиотеки с такого рода сервисами должны обеспечивать взаимодействие с существующими АБИС («электронными каталогами») на уровне импорта/экспорта метаданных (библиографических описаний), в том числе и с прикрепленными к ним файловыми ресурсами.

Архитектура таких библиотек должна быть ориентирована на работу в среде Интернет/Инtranet и допускать возможность «естественной» интеграции с другими информационными системами; в современных условиях это практически означает интеграцию на основе Веб-сервисов как в локальной сети, так и в распределенной среде.

В докладе рассматривается реализация изложенных выше подходов в рамках различных проектов, осуществленных и/или ведущихся (совместно с ООО «Константа») в рамках деятельности Архангельского областного центра повышения квалификации специалистов культуры (АОЦПК) в 2006–2011 гг., Архангельского краеведческого музея (АКМ) с 2011 г. по настоящее время, Российской ассоциации электронных библиотек (НП «ЭЛБИ») с 2011 г. по настоящее время. Использованы также идеи и подходы, реализованные нашими московскими партнерами из Библиотеки истории русской философии и культуры («Дом А.Ф.Лосева») и научно-справочной библиотеки Музеев Московского Кремля. Соответствующие электронные библиотеки созданы и развиваются в программной среде информационной системы T-Libra v.6.6.x (разработка ООО «Константа», г. Архангельск) [1, 4].

Во втором разделе доклада кратко описывается архитектура и основной функционал информационной системы T-Libra.

В третьем разделе дана более подробная характеристика полнотекстовых сервисов.

В четвертом разделе приводятся примеры использования этих сервисов для многоплановой поддержки аналитической и исследовательской деятельности, в том числе в рамках реализации некоторых направлений программы «Электронная Архангельская область».

В заключении содержится постановка вопроса о развитии полнотекстовых сервисов электронной библиотеки в направлении их интеллектуализации, а также о функциональной интеграции сервисов библиотеки с сервисами других информационных систем, создаваемыми в современной информационной среде.

2. Назначение, архитектура и функционал ИС T-Libra 6.6.x

2.1. Назначение и архитектура ИС T-Libra v.6.x

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. Она 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.

2.2. Основной функционал ИС T-Libra.

Для взаимодействия с существующими АБИС («электронными каталогами») реализован импорт метаданных (библиографической информации) в форматах ISO 2709 UniMarc / RusMarc в систему, в том числе вместе с присоединенными файлами полнотекстовых ресурсов (последние затем индексируются средствами T-Libra).

Импорт и индексация электронных полнотекстовых ресурсов в форматах RTF или PDF (графический с символьной текстовой подложкой или чисто символьный) для полнотекстового представления ресурса, сохраняющего базовое форматирование

исходных электронных текстов (разбивка на абзацы, постраничные и концевые сноски, вложенные изображения и таблицы, **жирность**, *курсив*, подчеркивание, р а з р я д к а); в разработке находится импорт форматов *.doc, *.html (в том числе html-массивов с сохранением гипертекстовых ссылок внутри массива), *.fb2.

Обеспечен импорт постраничного графического представления ресурса (например, в виде совокупности jpg-файлов), а также импорт текстовых, графических, аудио и видео файлов произвольных форматов для депозитарных представлений ресурса.

При импорте одновременно с индексацией полнотекстовых ресурсов происходит автоматическое пополнение многоязычного электронного словаря словоформ, возможно также и «ручное» редактирование словаря.

Другие возможности:

- обеспечивается поддержка Unicode при локализации интерфейса, при импорте текстов (создании словаря словоформ), при осуществлении поиска, при экспорте текстов из индексированной базы данных (отображении текста в окне браузера);
- поиск по интегрированному Каталогу (поддерживаются три формы поиска: однострочная, краткая, полная), в функционал одного из полей каталога встроен предметный мультирубрикатор, содержащий библиотечные классификаторы и рубрикаторы (УДК, ББК, ВАК, ГРНТИ и т.д.) в виде баз данных с собственными интерфейсами для их редактирования;
- использование для целей хранения и презентации пополняемого файлового хранилища (Депозитария) с файлами произвольного вида и формата и собственным настраиваемым рубрикатором ресурсов;
- гибкий тематизируемый многоязычный полнотекстовый поиск различного типа и вида с сортировкой, группировкой и различными формами презентации его результатов;
- возможности мультимодального расширения (графика, аудио, видео) функциональных возможностей ИС T-Libra 6.x., необходимое как для самой библиотеки, так и для взаимодействия и интеграции с другими информационными системами, модулями и оболочками (электронными коллекциями, электронными экспозициями, мультимедиа-энциклопедиями и т.д.);
- наличие функциональной подсистемы Личные ресурсы, предназначенная для создания и управления совокупностью персональных хранилищ информации, полученной с помощью сервисов ИС T-Libra и на ее ресурсной базе.

Хранилища информации, размещенные в этой подсистеме, используются также в качестве дополнительного удобного инструмента для формирова-

ния «корзины ресурсов» при осуществлении полнотекстового поиска.

3. Пользовательские сервисы электронной библиотеки

Электронные библиотеки, создаваемые на технологической основе ИС T-Libra, с самого начала задумывались нами (ООО «Константа») как *многофункциональные информационные среды*, используемые не только для предоставления доступа к ресурсам, но и для поддержки различных видов учебной, методической, научно-исследовательской, презентационной и организационно-управленческой деятельности, а также в качестве полигона для апробации новых информационных технологий и знакомства с ними специалистов библиотек, музеев, других учреждений сферы культуры, образования и науки.

В используемой нами версии электронной библиотеки имеются следующие типы полнотекстового поиска: **а) абзацно-ориентированный, б) частотно-ориентированный.** При этом абзацно-ориентированный поиск представлен разновидностями работы как в локальной, так и в распределенной среде.

3.1. Абзацно-ориентированный поиск

Он предназначен для поиска и презентации текста с точностью до отдельных авторских абзацев, содержащих заданную пользователем терминологическую структуру (тем самым эксплицируется «горизонтальный» микроконтекст, в котором в составе абзаца находятся искомые термины). Авторский абзац выбран в качестве естественной единицы смыслового членения текста. Обеспечивается поддержка нескольких видов и различных форм презентации результатов этого поиска:

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. Это поле функционирует аналогично стандартному поисковому полю в глобальных поисковых системах (Яндекс, Google и т.п.). 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).

Имеется возможность посмотреть, с этой же экранной страницы, соответствующий ресурс (статью, книгу и т.д.) в файловом виде; ресурс при этом может быть представлен в различных форматах: текстового документа, графического образа документа, сопровождающего документ аудио- или видеофайла.

Имеется также возможность оценки пользователем найденных абзацев, с пользовательскими комментариями к ним, и затем автоматизированной сборки выбранных абзацев с помощью опции **«Собрать тему»**. В результате пользователь получает файл, в котором собраны тематически ориентированные абзацы из различных документов электронной библиотеки, и указаны ресурсы (вместе с их библиографическими описаниями), откуда взяты эти абзацы. На клиентском (пользовательском) компьютере этот итоговый тематический файл может быть записан на переносимый носитель информации («флэшку») или распечатан.

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, например, из трех слоев сделать какие-либо два обязательными, а один — произвольно выбираемым при осуществлении запроса.

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 (например, «Сийское Евангелие», «факт истории», «Новодвинская крепость» и т.д.), до произвольной величины. Практически не имеет смысла задавать число больше 100 — поскольку в этом случае искомые слова в абзаце будут отстоять далеко друг от друга, и между ними с большей вероятностью не будет смысловой связи.

Замечание. Этот вид запроса может быть рассмотрен также как поиск с использованием *квазитезауруса*, создаваемого пользователем *ad hoc* («специально для данного случая»). Он, кстати говоря, может использоваться работниками библиотеки для обучения пользователей работе с полнотекстовыми ресурсами, в том числе составлению простых поисковых тезаурусов (например, формируемых в категориях: род-вид, часть-целое, концептуальные синонимы).

3.2. Частотно-ориентированный поиск

Предназначен для построения частотно-ранжированных списков терминов (существительных), и тем самым экспликации различных «вертикальных» макроконтекстов, неявно присутствующих в отдельном документе или их выбранной совокупности. Получающиеся таблицы списков терминов, с указанием абсолютного (в обычных числах) и относительного (в %, промилле) количества их встречаемости в тексте, мы называем «термиограммами» (по аналогии с «рентгенограммами»). Поиск может проводиться одновременно по 1, 2 или 3 корзинам ресурсов. Обеспечивается поддержка двух видов этого поиска и различных форм презентации его результатов:

абсолютный частотный, результатом которого является частотно-ранжированный список существительных, входящих в ресурсы области поиска и приведенных к нормальной форме (именительный падеж, единственное число).

относительный частотный, результатом которого является частотно-ранжированный список существительных, входящих только в те абзацы первичного текста, которые содержат заданный пользователем термин (тем самым список строится «относительно» этого термина).

Все термины, входящие в итоговую термиограмму, являются активными; «кликнув» по любому из них, можно выйти на *сформированный абзацно-ориентированный запрос* по данному термину, и эксплицировать его микроконтекст.

Эти виды частотного поиска могут использоваться для целей текстологического анализа документа; для выявления и описания предметной области документа; для составления списка ключевых слов; для сравнительного анализа предметных областей различных авторов или различных документов; для проверки текстов на плагиат и т.д.

Дополнительные возможности поиска и презентации его результатов.

Осуществляемая пользователем оценка релевантности результатов абзацно-ориентированного запроса, а также автоматическая сборка релевантных результатов запроса (т.е. тематической подборки абзацев) в один файл.

Отображение *графического оригинала* («имиджа») страницы текстового ресурса, содержащей найденный абзац текста (при условии предварительного постраничного импорта графического представления ресурса). Эта опция – шаг на пути объединения функций электронной библиотеки и электронного архива.

Сохранение запросов зарегистрированного пользователя для последующего формирования его «личного кабинета», хранимого на сервере.

4. Многоплановое использование полнотекстовых сервисов электронной библиотеки

4.1. Управленческая аналитика

Использование ресурсов и сервисов полнотекстовой библиотеки может оказать существенную поддержку развитию управленческой деятельности, в том числе осуществляемой в рамках развития *технологий электронного правительства*. Это касается по крайней мере трех из четырех основных направлений этого развития [9]: 1) повышение эффективности взаимодействия между различными ветвями государственной власти (*G2G, Government-to-Government*), 2) между государством и государственными служащими (*G2E, Government-to-Employees*), 3) между государственными органами и гражданами (*G2C, Government-to-Citizen*).

Замечание. Как справедливо отмечается в исследованиях [3], концепция электронного правительства, как она была сформулирована изначально, носит ярко выраженный нормативный характер: она операционализована, т.е. сведена к набору эмпирических индикаторов, которые могут быть собраны государственными органами. При этом социальный и культурный контексты, в котором функционируют управленческие органы, — и который, очевидно, должен учитываться в ресурсах и сервисах информационной среды — фактически игнорируется. На этот момент, ведущий к дисбалансу технологических и социокультурных ценностей, указывают и зарубежные исследователи [11].

Реальная практика продвижения программы в жизнь заставляет, однако, вносить коррективы в этот чрезмерно нормативный уклон. Характерно, что при моделировании информационной среды, поддерживающей задачи управления на всех уровнях власти, разработке онтологии и рубрикации предметной области «Государственное управление» приходится учитывать разнообразные официальные документы, статистику, административные регламенты, научные публикации, статьи СМИ, данные опросов и обследований, географические сведения и многие другие источники социальной, внутри- и внешнеполитической, экологической и т.д. тематики [10, с. 225-226]. Содержание соответствующих ресурсов, которые должны быть вовлечены в информационный оборот, далеко выходит за узкие индикативно-нормативные рамки концепции.

В рамках такого — расширенного — понимания концепции информационной среды «электронного правительства» мы считаем вполне уместным и важным использование ресурсов и сервисов электронной полнотекстовой библиотеки.

Приведем пример из нашей региональной практики.

Комитет по культуре Архангельской области с 1997 по 2010 г. издавал «Информационный бюллетень». 50 выпусков бюллетеня, каждый объемом от 25 до 30 уч.-изд. листов, содержат более 600 доку-

ментов (докладов, статей, аналитических обзоров, методических разработок, нормативных актов и т.п.) по актуальным вопросам региональной культурной политики. Его целевая аудитория — органы местного и государственного управления сферы культуры. Имеются как бумажная, так и электронная полнотекстовая версии этого бюллетеня.

Кроме того, ежегодно каждая подведомственная комитету организация (около 20 областных учреждений культуры) сдает содержательный отчет о своей деятельности, электронная полнотекстовая версия которого в среднем занимает объем в 50—60 страниц (около 1 Мб в формате Word-документа). Имеется большое количество других информационных ресурсов, накапливаемых в ходе деятельности органа управления культурой, но не входящих в систему административного документооборота.

Как правило, все такого рода ресурсы лежат в виде файлов на серверах администраций соответствующего уровня (иногда — на Веб-сайтах), и практически не участвуют в информационно-поисковом обороте.

На наш взгляд, место такого рода ресурсам — в специализированной (по ресурсам) электронной библиотеке с возможностями тематизируемого полнотекстового поиска. А самый этот поиск может использоваться для подготовки аналитических обзоров, справок, отчетов, при перспективном и текущем планировании и т.д.

Он может использоваться также для взаимодействия между различными государственными и муниципальными органами управления — например, департаментами культуры, образования, науки, туризма и т.п., поскольку сервисы универсальны, а ресурсы библиотеки относятся к самым различным областям деятельности граждан и организаций, и к тому же регулярно пополняются.

У нас есть небольшой опыт использования такого рода сервисов. В 2006–2011 гг. в Архангельском областном центре повышения квалификации была создана и функционировала библиотека с ресурсами по региональной культурной политике и вышеописанными сервисами полнотекстового поиска. Ее функционал использовался при подготовке различных информационных справок и обзоров.

Например, многослойный тематический запрос «Фестивали Архангельской области» (все три слоя актуальны, расстояние между терминами из разных слоев не более 8 слов) по всей полнотекстовой базе (насчитывавшей около 2000 документов) дает результат: 16 абзацев в 9 документах, удовлетворяющих параметрам запроса. Варьируя параметры запроса, можно легко получить другое количество релевантных абзацев/документов. Например, уменьшив расстояние между терминами, находящимися в разных слоях, до 1 слова (чтобы найти только те абзацы, где все три слова из запроса «Фестивали Архангельской области» примыкают друг к другу), мы получим в результате 2 абзаца в 1 документе. Можно одним «кликом» открыть полные тексты абзацев найденных документов. Используя кнопку

Контекст в левом меню, можно последовательно открыть от 1 до 3 абзацев до и после найденного, получив тем самым кластер из 7 абзацев. Можно, разумеется, открыть весь документ в виде файла (кнопка Ресурсы в электронной карточке документа).

Аналогично использовались поисковые сервисы для подготовки материалов по темам «Сохранение культурного наследия», «Информатизация сферы культуры», «Региональная культурная политика», «Реставрация икон и иконостасов», «Подготовка менеджеров культуры», «Финансирование муниципальных библиотек» и т.д.

4.2. Развитие информационных G2C -сервисов для граждан

Сервисы полнотекстовой библиотеки могут использоваться для информационной поддержки самых разных информационных систем и ресурсов, тем самым существенно повышая качество и увеличивая разнообразие информационных сервисов, предоставляемых гражданам.

Например, в 2010—2011 гг. в рамках реализации концепции «Электронная Архангельская область» была разработана региональная база данных «Памятники истории и культуры Архангельской области» (ООО «Константа» по заказу Инспекции по надзору за сохранностью памятников истории и культуры Архангельской области). БД «Памятники» содержит реестр сведений примерно о 2000 памятниках, с подробными описаниями и изображениями. Она предназначена одновременно и для служебного использования (представляет собой региональный Реестр памятников), и для общегражданского. Техническая поддержка этой БД обеспечивалась вначале службами областной администрации, а в настоящее время — областной научной библиотекой им. Н.А.Добролюбова.

В 2011–2012 гг. был начат проект по взаимодействию этой БД с электронной полнотекстовой библиотекой, содержащей около 2000 полнотекстовых ресурсов по актуальным вопросам культуры и региональной культурной политики. В настоящее время реализован автоматизированный запрос «от памятника — к библиотеке»: из интерфейса БД «Памятники» (со страницы с описанием конкретного памятника) одним «кликом» (например, от памятника «Русский гостиный двор») осуществляется тематически-ориентированный запрос, составленный предварительно экспертами, к полнотекстовым ресурсам библиотеки. Результатом запроса является список релевантных абзацев из различных документов (8 абзацев из 5 документов). Имеется также возможность доступа к файловым представлениям самих документов. При этом пользователь может далее самостоятельно работать как с БД «Памятники», так и с электронной полнотекстовой библиотекой по отдельности, не ограничиваясь только лишь экспертными запросами.

В ближайших планах участников проекта реализация более сложного автоматизированного запро-

са: «от библиотеки — к памятникам». При этом абзацы — результаты полнотекстового запроса по ресурсам библиотеки — автоматически будут связываться с релевантными памятниками: их названиями, изображениями и фрагментами описания; для этого потребуются автоматический семантический анализ найденных абзацев для последующего сопоставления их с описаниями памятников. Такой подход мы планируем распространить и на взаимодействие сервисов полнотекстовой библиотеки с другими информационными системами (электронными тематическими коллекциями, тематическими Веб-сайтами, открытыми электронными каталогами музейного и/или архивного фонда, электронными энциклопедиями и т.д.)

4.3. Тематизация материала для научно-образовательной деятельности

Сервисы полнотекстового поиска могут быть использованы также для тематических подборок материала при осуществлении научно-методической и научно-образовательной деятельности: подготовки лекций, экскурсий, методических пособий и т.д. Например, запрос по теме «Новодинская крепость» использовался при подготовке материала для различных мероприятий Архангельского краеведческого музея (АКМ); расстояние между терминами в двух различных слоях = 0 (термины в искомом абзаце примыкают друг к другу). По всем полнотекстовым ресурсам имевшимся на тот момент в распоряжении библиотеки АКМ (1756 документов) найдено 43 релевантных абзаца в 13 документах. Далее можно раскрыть каждый из найденных абзацев, оценить их с точки зрения пользователя, написать соответствующий комментарий, сохранить оценку (имеются соответствующие опции: Оценка абзаца, Комментарий, Сохранить оценку), затем вернуться на страницу с результатами запроса и нажать кнопку Собрать тему. Получим файл в формате *.html (сохраняемый средствами браузера) с тематической подборкой абзацев, который можно записать на флэшку или распечатать, и затем использовать в дальнейшей работе.

4.4. Экспликация смыслового контекста для гуманитарных исследований

Возможности полнотекстового поиска могут быть использованы для гуманитарных научных исследований — например, для экспликации терминологических полей (горизонтальных и вертикальных контекстов). Это сделано нами при экспериментальной разработке темы «*Экспликация концепта фактуальности в философии Н.А.Бердяева*».

С помощью относительного частотного поиска (относительно термина «факт») можно предварительно выявить смысловой макроконтекст, в котором существует концепт «факт» в философских текстах Н.А.Бердяева в целом или в тексте отдельного произведения. В частности, для исследуемых работ, содержащих термин «факт» (24 произведения) в

смысловую окрестность «факта» попадают (т.е. встречаются в абзацах, где имеется термин «факт»), в порядке убывания частотности, следующие термины: 1. «жизнь» (850 раз, 7,034 %), 2. «мир» (811 раз, 6,711 %), 3. «человек» (788 раз, 6,251 %), 4. «дух» (484 раза, 4,005 %), 5. «Бог» (483 раза, 3,997 %), 6. «сознание» (401 раз, 3,318 %), 7. «свобода» (377 раз, 3,120 %), 8. «факт» (366 раз, 3,029 %), 9. «философия» (324 раза, 2,681 %), 10. «история» (311 раз, 2,574 %).

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

Можно использовать многослойный запрос. В один слой ввести термин «факт», в два других — термины из бердяевской терминограммы (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 полнотекстового запроса (эксплицируются микроконтексты концепта фактуальности, связанные с темами:

- «гуманизм и факт рабства человека у природного мира»;
- «факт самосознания человека и трансцендентность сознания» и т.д.).

Выбирая другие термины из терминограмм 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, терминограммам Лосева и Канта и соответствующим абзацно-ориентированным запросам по их произведениям. Это дает, в свою очередь, 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 смысловой («вертикальный») макроконтекст, в котором существует концепт «одино-

чество» 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, 8.175 %), 2. «человек» (384 раза, 6.679 %), 3. «жизнь» (334 раза, 5.809 %), **4. «одиночество» (308 раз, 5.357 %)**, 5. «личность» (296 раз, 5.148 %), 6. «общение» (187 раз, 3.253 %), 7. «люди» (175 раз, 3.044 %), 8. «познание» (170 раз, 2.957 %), 9. «общество» (169 раз, 2.939 %), 10. «существование» (156 раз, 2.713 %).

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 многослойный запрос (см. выше). 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 терминологии 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.

Проблема одиночества представляется мне основной философской проблемой, с ней связаны проблемы «я», личности, общества, общения, познания.

Включая другие различные сочетания терминов 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 выше обозначенной 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, терминологиям М.К. Мамардашвили и С.Л. Франка и соответствующим абзацно-ориентированным запросам по их произведениям.

Предварительные результаты этого исследования были представлены также на одном из заседаний Круглого стола конференции Телематика-2013 (СПб.: НИУ ИТМО, 26 июня 2013 г.)

Заключение

Дальнейшее развитие подобных многофункциональных информационных сервисов мы связываем с *интеллектуализацией* поисковых и презентационных возможностей электронной полнотекстовой библиотеки: семантическим картированием, включением тезаурусов в полнотекстовый поиск, разработкой комбинированных и каскадных запросов и т.д.

Это направление развития может и должно быть дополнено:

- интеграцией полнотекстовых ресурсов и сервисов в составе распределенной информационной среды регионального и/или отраслевого уровня [6];
- функциональной интеграцией сервисов электронной библиотеки с сервисами электрон-

- ных коллекций, динамических веб-сайтов, энциклопедий и других баз данных [5];
- созданием различных виртуальных образовательных и исследовательских инфраструктур, при этом организационно-психологические факторы обеспечения эффективного взаимодействия оказываются не менее важными, чем технологические [8].

Литература

- [1] Демоверсия электронной библиотеки ООО «Константа». URL: <http://demo.tlibra.ru> (дата обращения: 21.08.2013).
- [2] Дерновой Г. Семантический анализ на РОМИП-2009 URL: http://romip.ru/romip2003/7_dernovoy.pdf (дата обращения: 20.08.2013)
- [3] Дьякова Е.Г. Электронное правительство как нормативная концепция: пределы применимости // Труды XV Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2012), Санкт-Петербург, Россия, 2012. С. 280—283.
- [4] Ляпин С.Х., Куковякин А.В. Многоязычный поиск в электронной библиотеке и его реализация в ИС T-Libra 6.x // XI Международная конференция EVA-Moscow 2008, Москва, 1—5 дек. 2008 г. URL: http://conf.cpic.ru/upload/eva2008/reports/doklad_1389.doc (дата обращения: 21.08.2013).
- [5] Ляпин С.Х., Куковякин А.В. Функциональная интеграция библиотеки, коллекций, энциклопедии и веб-сайта в распределенной информационной среде // Труды XI Всерос. объедин. конф. «Интернет и современное общество» (IMS-2008). 28—30 окт. 2008 г., Санкт-Петербург, СПбГУ. СПб.: СПбГУ, 2008. С. 72—75. URL: http://conf.infosoc.ru/2008/pdf_HI/Lyapin&Kukovyakin.pdf.
- [6] Ляпин С.Х., Куковякин А.В. Электронная библиотека с распределенным полнотекстовым поиском: на пути к функциональной интеграции ресурсов и сервисов. // Доклад на XX междунар. конф. «Крым-2013» (презентация). URL: <http://www.aselibrary.ru/conference/conference43/conference434199/> (дата обращения: 21.08.2013).
- [7] Смирнов И.В., Соченков И.В., Тихомиров И.А. Система интеллектуального поиска и анализа информации «Exactus» на РОМИП-2009. URL: http://romip.ru/romip2009/03_exactus.pdf. (Дата обращения: 20.08.2013).
- [8] Шварцман М.Е. Электронная библиотека как основа виртуальной исследовательской инфраструктуры. URL: <http://www.aselibrary.ru/blogs/archives/1209/> (дата обращения: 20.08.2013).
- [9] Электронное правительство и качество государственного управления. URL: <http://open-gov.ru/2012/11/04/elektronnoe-pravitelstvo-i-kachestvo-gosupravleniya-2/> (Дата обращения: 21.08.2013).
- [10] Юдина Т.Н., Богомолова А.В. УИС РОССИЯ: онтология предметной области «государствен-

ное управление» // Труды XIV Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2011). СПб.: 2011. С. 225—230.

- [11] Heeks R., Santos R. Understanding Adoption of e-Government: Principals, Agents and Institutional Dualism. URL: http://www.sed.manchester.ac.uk/idpm/research/publications/wp/igovernment/igov_wp19.htm.

Digital full-text library for support of analytical & research working

S. Lyapin, A. Kukovyakin

The report discusses the use of the services of multifunctional digital full-text library, created on the basis of the information system of the T-Libra 6.6.x («Constanta» Ltd., Arkhangelsk, Russia), to support analytical and research activities. Examples of such use are given: for support of the management activity in the sphere of culture (analytics of reports, thematic analysis of information materials); activities of the Museum (the information-analytical support of exhibitions and expositions); humanitarian researches in the sphere of philosophy (explication of semantic contexts of Russian philosophy).