

КАК ПРОЙТИ В РАСПРЕДЕЛЕННУЮ БИБЛИОТЕКУ?

HOW TO PASS IN THE DISTRIBUTED LIBRARY?

S. Lyapin

Annotation

One of approaches locates in article to the organization service-oriented distributed information environment of a middle (branch and/or regional) level; in this context the question of creation of the distributed inter-museum electronic full-text library is discussed. The main functional of the electronic full-text library developed now in Arkhangelsk museum of local history on the basis of information system T-Libra (development of Constanta Ltd., Arkhangelsk) is considered. In this regard the attention to the question of functioning of the distributed library environment based on principles of the federal (decentralized) organization, and possessing possibilities of the distributed full-text search is brought. Experiment on realization of such search, carried out in April, 2011 within conference "Museum libraries in the modern world", passing on the basis of Museums of the Moscow Kremlin is described. The description of the project preparing for realization on creation in 2013–2015 of the distributed regional inter-museum electronic library in the Arkhangelsk region is given. Questions of use of offered information services for support of primary activity of a museum are discussed.

Keywords: distributed full-text search, resources & services integration, museum digital library.

Ляпин Сергей Хамзеевич

Кандидат философских наук
Архангельский краеведческий музей;
ООО "Константа", Архангельск

Аннотация

В статье обосновывается один из подходов к организации сервис-ориентированной распределенной информационной среды среднего (отраслевого и/или регионального) уровня; в этом контексте обсуждается вопрос о создании распределенной межмузейной электронной полнотекстовой библиотеки. Рассматривается основной функционал электронной полнотекстовой библиотеки, развиваемой в настоящее время в Архангельском краеведческом музее на базе информационной системы T-Libra (разработка ООО "Константа", Архангельск). В этой связи ставится вопрос о функционировании распределенной библиотечной среды, основанной на принципах федеративного (децентрализованного) устройства, и обладающей возможностями распределенного полнотекстового поиска. Описан эксперимент по реализации такого поиска, проведенный в апреле 2011 года в рамках конференции "Музейные библиотеки в современном мире", проходившей на базе Музеев Московского Кремля. Дано описание готовящегося к реализации проекта по созданию в 2013–2015 гг. распределенной региональной межмузейной электронной библиотеки в Архангельской области. Обсуждаются вопросы использования предлагаемых информационных сервисов для поддержки основной деятельности музея.

Ключевые слова:

Распределенный полнотекстовый поиск, интеграция ресурсов и сервисов, музейная электронная библиотека.

Введение. Сервис-ориентированная распределенная информационная среда сферы культуры: место и роль электронной библиотеки

Современный этап развития информационных систем в сферах образования, культуры и науки характеризуется, на наш взгляд, следующими важными моментами.

Во-первых, переходом от элементарных информационных сервисов (поиск по каталогу + доступ к электронным ресурсам в их файловом виде, представленными, как правило, в графическом формате), к продвинутым сервисам гибкого тематизируемого полнотекстового поиска по полнотекстовым и полносодержательным ресурсам и различным формам презентации результатов этого поиска.

Во-вторых, интеграцией ресурсов и сервисов как в рамках информационной среды конкретной организации (библиотеки, музея, образовательного или научного уч-

реждения), так и в рамках более широкой информационной среды.

В-третьих, переходом от локальных решений к сервисам (в том числе интегрированным) в распределенной информационной среде. При этом технологическая организация распределенной информационной среды может быть весьма различной, с большей или меньшей степенью централизации управлением информационными сервисами.

В реализации всех вышеназванных тенденций исключительно важную роль играют электронные библиотеки, – наиболее универсальные по ресурсам и сервисам информационные системы.

В статье обосновывается один из возможных подходов к организации сервис-ориентированной распределенной информационной среды среднего (отраслевого и/или регионального) уровня. Речь идет о создании рас-

пределенной межмузейной электронной полнотекстовой библиотеки и ее использовании для информационной поддержки основной деятельности музеев: экспозиционно-выставочной, экскурсионной, научно-методической, атрибуционной, просветительской.

Ведь сегодня "выигрывает тот, кто полнее других использует новые возможности" [7, С. 29].

1. Полнотекстовый поиск в электронной библиотеке. Краткое описание основного функционала (для информационной системы T-Libra v. 6.x)

Работы по созданию многофункциональной электронной полнотекстовой библиотеки были начаты нами еще в середине 1990-х гг. (в эпоху полного доминирования электронных каталогов) в рамках деятельности научно-информационного центра Поморского государственного университета им. М.В. Ломоносова и ООО "Константа", и продолжены в рамках различных проектов. Результаты исследований и разработок докладывались на крупнейших российских и международных конференциях 1998–2012 гг. (см. напр.: [2; 3; 4]).

В статье описываются возможности такого подхода на примере информационной системы T-Libra, к настоящему времени работающей в регулярном режиме в нескольких организациях России, в том числе в Музеях Московского Кремля, библиотеке истории русской философии и культуры "Дом А.Ф. Лосева" (Москва), в Архангельском краеведческом музее.

ИС T-Libra функционируют в клиент-серверной Интернет/Интранет архитектуре. На стороне пользователя предполагается лишь наличие Интернет-браузера и стандартных прикладных программ по работе с файловыми ресурсами. На стороне сервера – операционная система Windows, СУБД MySQL, Веб-сервер Apache.

В используемой нами версии электронной библиотеки [1] 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.

Абзацно-ориентированный поиск 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 смысло-

вого 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 (Yandex, 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 электронной библиотеке в различных форматах: текстового документа, графического образа документа, сопровождающего документ аудио- или видео-файла и т.д. 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 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; внутри слоя – операция логического объединения 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", или "факт истории" 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 большой вероятностью не будет смысловой связи.

Этот вид запроса может быть рассмотрен также как поиск с использованием квази-тезауруса, создаваемого пользователем ad hoc ("специально для данного случая").

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 \square , промилле, pro mille) количества их встречаемости в тексте, мы называем "терминограммами" (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 таблицу – "терминограмму" – являются активными; "кликнув" по любому из них, можно выйти на уже сформированный абзачно-ориентированный запрос по данному термину, и экспликовать 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 и т.д.

2. Распределенный полнотекстовый поиск. Эксперимент по его реализации.

Приступая в конце 2010 г. к разработке распределенного полнотекстового поиска, мы (ООО "Константа") обнаружили, что не существует международных стандартов на эту тему, соответственно – нет и готовых протоколов осуществления такого поиска в среде Интернет (стандарты и протоколы такого рода есть только для поиска по каталогу). В этой ситуации была поставлена и решена задача для частного случая: распределенного полнотекстового поиска для однородных информационных систем, находящихся в различных точках Интернет-пространства. В качестве экспериментальной информационной системы была выбрана наша собственная разработка – ИС T-Libra v.6.x, установленная к тому времени в нескольких организациях Архангельска и Москвы.

5 апреля 2011 г. в рамках ежегодной конференции "Музейные библиотеки в современном мире", проходившей в Музеях Московского Кремля, был осуществлен (впервые в России) эксперимент по реализации распределенного полнотекстового поиска. В нем участвовали электронные библиотеки 5 организаций: 2 в Архангельске (Архангельский областной центр повышения квалификации специалистов культуры и ООО "Константа") и 3 – в Москве (Музеи Московского Кремля, библиотека истории русской философии и культуры "Дом А.Ф. Лосева", библиотека Государственного исторического музея).

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 20 секунд (параметр регулируется пользователем при формировании запроса), этого оказалось достаточно для успешного ответа всех библиотек. Были осуществлены оба варианта абзачно-ориентированного поиска (однослойный и многослойный 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 (8 апреля 2011 г.) в Российской ассоциации электронных библиотек [6].

3. О создании межмузейной распределенной электронной библиотеки в Архангельской области в 2013-2015 гг.

24 октября 2012 г. на региональном музейном семинаре "Музей в современном информационном пространстве", проходившем в Архангельском краеведческом му-

зее (АКМ), была продемонстрирована работа электронной полнотекстовой библиотеки АКМ, в том числе и в режиме распределенного полнотекстового поиска.

В этой связи был анонсирован проект "Создание региональной распределенной межмузейной электронной полнотекстовой библиотеки", включенный в государственную программу развития культуры Архангельской области в 2013–2015 гг.

На первом этапе (2013 г.) на основе информационной системы Т Libra будет создана распределенная информационная среда, включающая электронную библиотеку АКМ и двух районных музеев Архангельской области; на последующих – включение в распределенную среду еще 8 районных музеев.

Реализация этого проекта позволит:

а) создать и поэтапно развивать электронную библиотеку с полнотекстовым поиском в каждом из музеев, использовать ее для информационной поддержки всех основных видов музейной деятельности (экспозиционно-выставочной, экскурсионной, научно-исследовательской, научно-методической, атрибуции музейных фондов);

б) предложить пользователю ("посетителю") качественно новый вид и уровень информационного сервиса.

в) использовать библиотеку для создания электронного научного архива в каждом из музеев, что позволит ввести в оборот уникальные документы и материалы, собранные в музейных архивах (только в архиве АКМ находится около 2500 папок с материалами археологических и этнографических экспедиций);

г) функционально объединить библиотечные ресурсы и доступные для пользователей сервисы всех участвующих в проекте музеев;

д) существенно расширить пользовательскую аудиторию каждого из музеев, поскольку пользователь ресурсов и сервисов библиотеки одного из музеев автоматически становится пользователем ресурсов и сервисов всех других музеев – участников распределенной среды; при этом активность пользователя автоматически фиксируется подсистемой сбора статистических данных электронной библиотеки.

Планируется проведение обучающих семинаров для участников проекта, вначале на базе АКМ, а затем и в рамках выездных мероприятий. Координацией деятельности по созданию цифровых ресурсов, по их включению в сервисы электронной межмузейной библиотеки, разработкой методических материалов займется областная краеведческий музей. Рассматривается возможность создания на этой основе Консорциума музейных библио-

тек с постепенным расширением числа его участников, причем не только из Архангельской области.

4. Место и роль электронной полнотекстовой библиотеки в информационной поддержке основной деятельности музея

Электронная библиотека с возможностями гибкого тематизируемого полнотекстового поиска, "достающего" до авторских абзацев и других произвольных фрагментов текста создает принципиально новые возможности для информационной поддержки основной музейной деятельности. Тем более это относится к распределенной полнотекстовой библиотеке.

В этой связи на вышеупомянутом семинаре обсуждались вопросы:

◆ Взаимодействие библиотеки и научного архива музея. На первых порах предполагается создание электронного научного архива в составе электронной полнотекстовой библиотеки. Архивные материалы могут при этом описываться в рамках существующего библиографического каталога (поля "автор", "заглавие", "ключевые слова", "аннотация", "год издания", возможно использование библиографических классификаторов и рубрикаторов); кроме того, можно попробовать использовать некоторые поля стандарта RUSMARC (например, блок 900-тых полей) для отображения специфически-архивных моментов ("фонд", "опись", "дело").

Полнотекстовые ресурсы могут создаваться в комбинированном виде: графические образы ("имиджи") всего документа плюс его избранные фрагменты (дайджесты) в текстовом (символьном) виде, которые и включаются в механизмы индексации и полнотекстового поиска. Их взаимодействие обеспечивается подсистемами визуализации документов (связь найденного абзаца с графической страницей, на которой он находится в первичном ресурсе).

Библиотека и каталог музейного фонда. Мы предполагаем проведение переговоров с разработчиками электронных каталогов музейного фонда (фактически – это две организации в России: ГИВЦ Минкультуры РФ и ЗАО "КАМИС") об организации эксперимента по осуществлению интеграции ресурсов и сервисов электронного каталога и музейной электронной полнотекстовой библиотеки.

Библиотека и издательская деятельность музея. Электронная библиотека предоставляет прекрасную возможность осуществлять публикацию материалов, создаваемых сотрудниками музея – от официальных годовых отчетов до отдельных статей, сборников, материалов конференций и семинаров, методических разработок и т.п.

Библиотека и выставки (экспозиции).

Основные направления использования ресурсов и сервисов электронной библиотеки для этих важнейших видов музейной деятельности:

а) отбор и подготовка материала для музейных этикеток и аналитических описаний экспонатов;

б) использование библиотеки в качестве дополнительных сервисов для посетителей экспозиций и выставок музея (доступ для Интернет-пользователей через Веб-сайт музея; оборудование в залах музея рабочих мест для посетителей музея с прямым доступом к корпоративной библиотеке музея);

в) натурные и электронные экспонаты с полнотекстовой поддержкой;

г) взаимодействие электронной библиотеки с электронными тематическими коллекциями, выставками и экспозициями.

Заключение.

Актуальные вопросы развития распределенных информационных систем сферы культуры

Распределенная информационная среда сферы культуры не ограничивается, разумеется, только электронной полнотекстовой библиотекой [5]. Так, в Архангельской

области в 2010–2011 гг. Инспекцией по охране памятников, ООО "Константа" и Архангельским областным центром повышения квалификации специалистов культуры (АОЦПК) реализован проект по взаимодействию в распределенной среде региональной базы данных "Памятники истории и культуры Архангельской области" (описания и изображения около 2000 памятников) и тематической электронной полнотекстовой библиотеки.

К 300-летию со дня рождения М.В. Ломоносова (2011 г.) был осуществлен экспериментальный проект по созданию интегрированной распределенной информационной среды "Пространство Ломоносова", обеспечивающей взаимодействие ресурсов и сервисов виртуального атласа "Земля Ломоносова", созданного с использованием технологии Google Earth (изображение и описание около 50 локаций, связанных с жизнедеятельностью М.В. Ломоносова), и научно-мемориальной полнотекстовой библиотеки Ломоносова, разработанной на основе технологии T Libra. Организаторы и разработчики проекта: Институт культурного и природного наследия им. Д.С. Лихачева (Москва), ООО "Константа" и АОЦПК (Архангельск).

Продвинутые поисковые и презентационные сервисы по самим ресурсам (а не только по каталогам), функциональная интеграция ресурсов и сервисов на основе открытых распределенных информационных систем различного уровня – таков, на наш взгляд, магистральный путь развития современного информационного пространства сферы культуры.

ЛИТЕРАТУРА

1. Информационная система T-Libra 6.6. [Электронный ресурс] // URL: <http://demo.tlibra.ru>
2. Ляпин С.Х., Куковякин А.В. Многофункциональная электронная библиотека на основе ИС T-Libra v.6.x для поддержки музейных исследований // Двенадцатая ежегодная конференция АДИТ-2008 "Культурное наследие регионов России в мировом информационном пространстве". 2–6 июня 2008 г. Россия, г. Пермь. [Электронный ресурс] // URL: <http://www.adit.ru/rus/conference/adit2008/papers/paper.asp?nomer=37>
3. 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 декабря 2008 г. [Электронный ресурс] // URL: http://conf.cpic.ru/upload/eva2008/reports/doklad_1389.doc
4. Ляпин С.Х., Куковякин А.В. Распределенная электронная библиотека: новые подходы к интеграции ресурсов и сервисов. Тезисы доклада на конференции EVA-2011 Москва, Российская государственная библиотека, 30 ноября 2011 года. [Электронный ресурс] // URL: <https://eva.rsl.ru/ru/2011/report/list/973/all/82?page=3>.
5. Ляпин С.Х., Куковякин А.В. Функциональная интеграция библиотеки, коллекций, энциклопедии и веб-сайта в распределенной информационной среде // Труды XI Всеросс. объединенной конференции "Интернет и современное общество" (IMS-2008). – 28–30 окт. 2008 г., Санкт-Петербург, СПбГУ. – Изд. Института искусств Факультета филологии и искусств СПбГУ, 2008. С. 72–75. [Электронный ресурс] // URL: http://old.conf.infosoc.ru/2008/pdf_HI/Lyapin&Kukovyakin.pdf
6. Презентация проекта "Распределенная межмузейная электронная библиотека". Видеоконференция Российской ассоциации электронных библиотек (НП "ЭЛБИ"), 08 апреля 2011 г. [Электронный ресурс] // URL: <http://www.aselibrary.ru/conference/conference43/conference432039>
7. Путин В.В. О наших экономических задачах // Государственная служба. Вестник Координационного Совета по кадровым вопросам, государственным наградам и государственной службе при полномочном представителе Президента Российской Федерации в Северо-Западном федеральном округе. 2012. № 1. С. 29–44.