

Адрес статьи / To link this article: <http://cat.ifmo.ru/ru/2017/v2-i1/98>

Пространство Ломоносова: опыт функциональной интеграции виртуального атласа и полнотекстовой библиотеки

С.Х. Ляпин¹, А.В. Куковякин²

¹ Университет ИТМО, Архангельский краеведческий музей, Россия

lyapins@yandex.ru

² Архангельский краеведческий музей, ООО «Константа», Россия

magicmagus@yandex.ru

Аннотация. Описывается опыт создания функционально интегрированной распределенной мультимодальной информационной среды, включающей в себя: а) электронную библиотеку с поиском по полнотекстовым ресурсам, относящимся к деятельности М.В.Ломоносова; б) виртуальный атлас географических мест в России и за рубежом, связанных с именем М.В.Ломоносова, в) функциональное взаимодействие сервисов виртуального атласа с сервисами электронной полнотекстовой библиотеки. Виртуальный атлас создается на основе технологии Google Earth; на виртуальном глобусе определяются места, связанные с жизнью и деятельностью М.В.Ломоносова (около 100 локаций), создаются тематические слои, активные объекты которых взаимодействуют с тематическим поиском по электронной полнотекстовой библиотеке. Публикация подготовлена при поддержке гранта РГНФ № 14-03-12017.

Ключевые слова: сервисы полнотекстового поиска, мультимодальное взаимодействие, сервис-ориентированная архитектура, виртуальный атлас, функциональная интеграция сервисов, абзацно-ориентированный запрос.

1. Введение. О методологии и архитектуре

Растущее количество и увеличивающееся разнообразие информационных систем, существующих в современном информационном пространстве и актуально используемых в деятельности конкретных организаций — библиотек, музеев, университетов — ставит в повестку дня вопрос об их функциональной интеграции, как в пределах локальных сетей, так и в различных конфигурациях внешней информационной среды.

При этом нам кажется важным сделать акцент не просто на интеграции ресурсов в рамках той или иной технической платформы, а на интеграции поисковых (и презентационных) сервисов. Мы делаем это в рамках методологии мультимодального взаимодействия и разговорного анализа [1], а с точки зрения стиля программного проектирования — в рамках парадигмы сервис-ориентированной распределенной среды.

Методология мультимодального взаимодействия (и разговорного анализа) предполагает рассмотрение любого «действия» человека — в том числе, например, пользователя информационной системы — с точки зрения взаимодействия различных «модальностей» информации: текста, графики (статических образов), аудио (речи), видео (динамических образов) и т.д. для организации комплексного смыслового поля, в котором находится человек в реальном жизненном пространстве, а пользователь, соответственно — в информационном пространстве.

Парадигма сервис-ориентированной архитектуры (COA, SOA англ. service-oriented architecture) — это модульный подход к разработке программного обеспечения, основанный на использовании распределённых, слабо связанных компонентов, оснащённых стандартными интерфейсами для взаимодействия по стандартным протоколам [2].

Мы в наших разработках ориентируемся на веб-сервисы и веб-протоколы.

В докладе рассматривается организация мультимодальной информационной среды, включающей в себя виртуальный гипермедиа атлас с локациями, относящимися к жизни и деятельности М.В.Ломоносова, и электронную полнотекстовую библиотеку, включающую в себя ресурсы, относящиеся к деятельности М.В.Ломоносова (но не только). Работы проводятся в интересах Архангельского краеведческого музея, и направлены на поэтапное создание современной информационной среды для поддержки основной деятельности музея (экспозиции, выставки, экскурсии), а также на создание полнофункционального виртуального музея.

2. Виртуальный атлас на основе технологии Google Earth

Функциональную часть распределенной информационной среды «Пространство Ломоносова» составляет виртуальный гипермедиа атлас «Земля Ломоносова». Он создается на основе технологии Google Earth [3] (бесплатная версия) и содержит около 100 локаций («метки» на глобусе), связанных с именем и деятельностью М.В. Ломоносова.

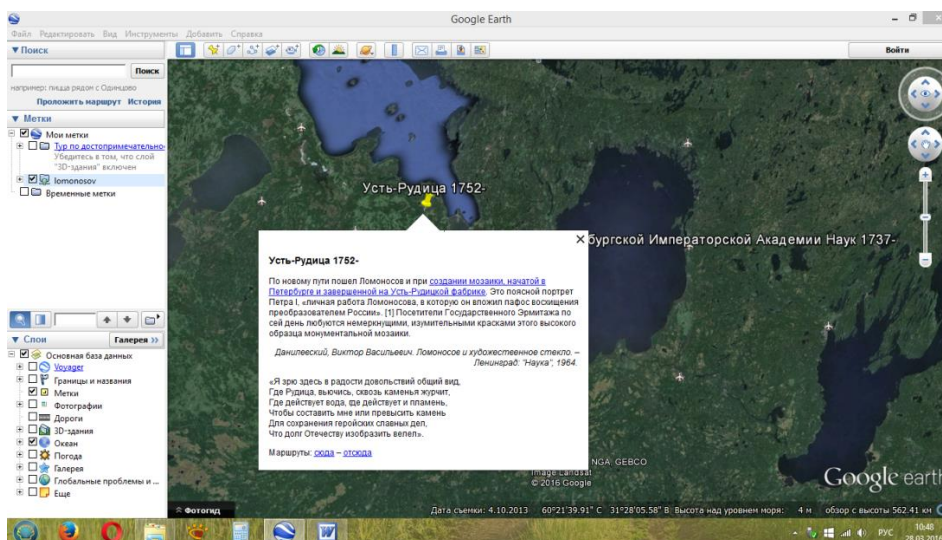


Рис.1 Скриншот страницы виртуального атласа Google Earth с локацией «Усть-Рудица, 1752» и этикеткой локации, в которой выделена ссылка на автоматический запрос к библиотеке.

Эти локации содержат краткое описание содержания локации (в терминах музейного дела — это «этикетки предмета»), а также ссылку на автоматический запрос к полнотекстовой библиотеке с ресурсами по ломоносовской тематике.

На глобусе могут быть размещены также ссылки на другие элементы создаваемой информационной среды: электронные экспозиции и выставки, тематические электронные коллекции с поиском и навигацией (например, в разработке находится мультимодальная коллекция «Концепты Ломоносова», посвященная особенностям понятийно-концептуального аппарата в научном творчестве М.В.Ломоносова), другие специализированные базы данных. Их

взаимодействие между собой каждый раз определяется и организуется сообразно поставленным задачам.

3. Полнотекстовая библиотека на основе информационной системы T-Libra

Важнейшей частью создаваемой информационной среды является электронная библиотека с возможностями гибкого тематического полнотекстового поиска.

Физически библиотека может находиться на любом компьютере (сервере), имеющем связь с Интернет, а логически — в составе децентрализованной распределенной информационной среды.

Используемая нами в различных проектах информационная система T-Libra, предназначенная для создания многофункциональных электронных полнотекстовых библиотек (разработка ООО «Константа», Архангельск), функционирует 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 (или аналогичная по функционалу), Веб-сервер Apache, сервер приложения (ООО «Константа»). Вся бизнес-логика вынесена в сервер приложения [4].

Примечание: T-Libra может функционировать также в *распределенной однородной информационной среде*: на серверах участников этой среды должна быть установлена ИС T-Libra; пользователь запускает сквозной запрос с любого из серверов, и получает результаты от всех тех серверов, к которым разрешен доступ. [5] Тем самым реализована модель децентрализованной среды под управлением пользовательского браузера, обращающегося к множеству независимых серверов, находящихся в ведении различных организаций (архитектура «машины метапоиска», Metasearch engine [6]. Этот подход вполне укладывается в парадигму распределенных информационных систем [7], достаточен для целей полнотекстового поиска и позволяет не рассматривать более сложные варианты.

3.1. Сервисы полнотекстового поиска

В текущей 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.

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 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. 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 вертикальном 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 (важно для архивных ресурсов), сопровождающего документ аудио- или видеофайла (важно для организации электронных выставок и коллекций).

Имеется также возможность оценки пользователем найденных абзацев, 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 содержит функционал дополнительной 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. Например, в первом слое вводим термин «Ломоносов», во втором — термин «Усть-Рудица», в третьем — термин «мозаика». Тем самым в структуре запроса тематика «Ломоносов» специализирована (аспектуализирована) в связи с «Усть-Рудицей» и «мозаикой».

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»), до произвольной величины.

В запросе «от глобуса — к библиотеке» (составленном экспертами) используется, как правило, *многослойный абзацно-ориентированный запрос*.

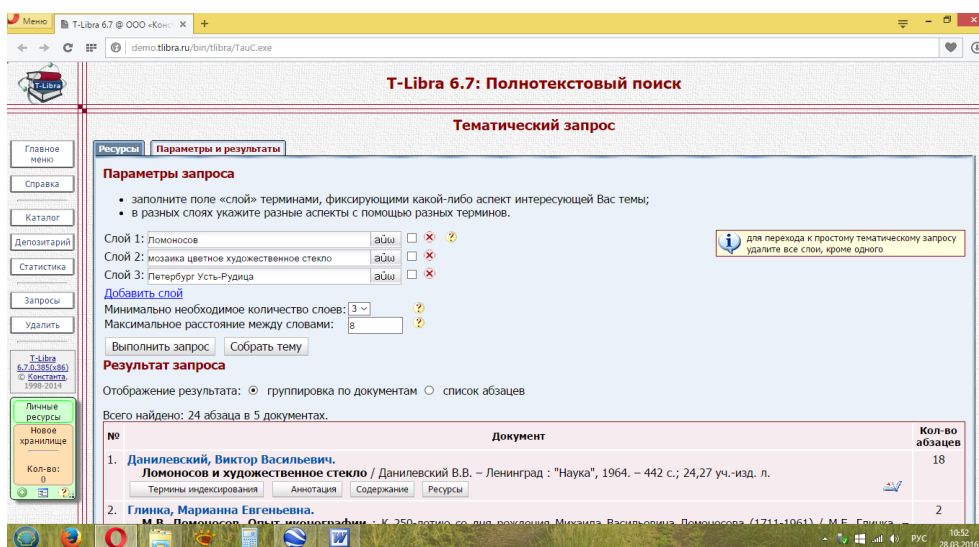


Рис.2. Скриншот страницы с результатом автоматического запроса «от локации Усть-Рудица — к полнотекстовой библиотеке».

В результате трехслойного запроса (слой 1 «Ломоносов», слой 2 «мозаика цветное художественное стекло», Слой 3 «Петербург Усть-Рудица»; расстояние между словами в разных слоях не более 8 слов) — см. выше рис. 2 — найдено **24 релевантных абзаца в 5 документах** (по базе объемом около 3500 документов). Термины внутри одного слоя подчиняются операции *логического объединения*, термины в разных слоях – операции *логического пересечения*.

Каждый из абзацев одним «кликом» может быть раскрыт до своего полного вида (см. ниже рис.3).

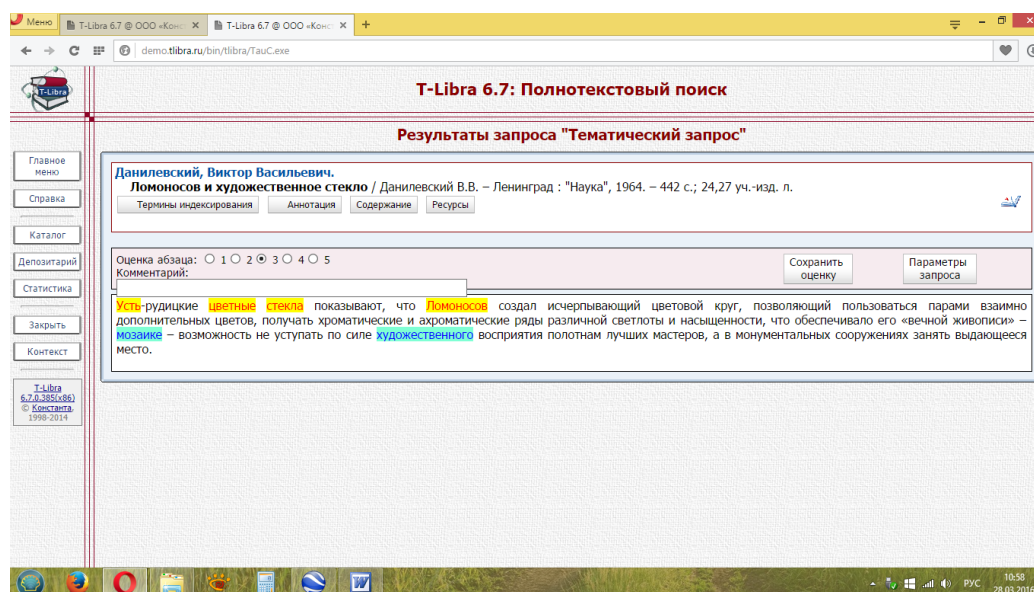


Рис. 3 Раскрыт один из найденных абзацев (из книги В.В.Данилевского «Ломоносов и художественное стекло»). Термины запроса, удовлетворяющие условиям его фокусировки, подкрашены красным цветом текста на **желтом фоне**.

Далее можно воспользоваться сервисами самой библиотеки. Например, актуализировать контекст, в котором находится найденный абзац (кнопка «контекст» в меню) — можно открыть абзац до и абзац после абзаца-результата запроса, и в целом кластер из 7 абзацев: три абзаца до и три абзаца после найденного, плюс сам абзац результата запроса.

Можно также организовать самые разные запросы по интересующей пользователя теме (уже безотносительно к гипермедиа атласу), затем воспользоваться опцией «собрать тему» и получить файл с отображенными абзацами и ссылками на ресурсы, из которых они взяты. Наконец, можно открыть сами файлы ресурсов в различных форматах (кнопка «ресурсы» в электронной карточке соответствующего документа).

Таким образом, поиск «от виртуального глобуса — к электронной библиотеке» позволяет в интерактивном режиме связать локации на глобусе с релевантными тематическими абзацами электронной библиотеки, тем самым актуализовать культурно-исторический контекст соответствующих событий и имен, связанных с деятельностью М.В.Ломоносова.

В настоящее время ведутся работы по решению обратной, более сложной задачи — автоматизированному поиску «от библиотеки — к виртуальному глобусу»: от найденного в библиотеке абзаца одним кликом получить те локации на глобусе, которые тематически релевантны содержанию абзаца. Это предполагает включение в состав поисковых сервисов более глубокого семантического анализа текста/контекста, «достающего» до смысловых единиц информации.

4. Заключение

В целом информационную среду в составе [виртуальный гипермедиа атлас + полнотекстовая библиотека] мы рассматриваем как один из вариантов построения **унифицированного мультимодального портала** для входа в информационную среду «Пространство Ломоносова».

За счет развития ресурсов и сервисов гипермедиа атласа, присоединения к ним различных тематических коллекций, функциональной интеграции этих элементов в составе атласа, а также двусторонней интеграции между атласом и полнотекстовой библиотекой мы получаем возможность многопланового расширения и углубления информационной среды современного музея.

Сказанное относится, разумеется, не только к ломоносовской тематике.

В перспективных планах музея — применение методологии мультимодального взаимодействия и технологии сервис-ориентированных систем для таких важных комплексных тем, как «Наследие Архангельского епархиального древлехранилища» (с него началась коллекция Архангельского краеведческого музея), «Словесность Русского Севера: тексты, контексты, образы», «Северная икона в музейном информационном пространстве», «Народная культура Русского Севера», «Новодвинская крепость», «Освоение Арктики» и т.д.

Примечание. Трехлетний опыт работы в рамках изложенной выше программной среды показал, что ряд функций, которые хотелось бы реализовать в рамках унифицированного мультимодального портала, не вполне соответствует функционалу Google Earth. В частности, не может быть эффективно решена задача создания интерактивных виртуальных экскурсий с изменяемой пользователем траекторией движения по музейным артефактам. В связи с этим в настоящее время ведутся разработки интегрированной программной среды (коллекции + библиотека + интерактивная виртуальная экскурсия), полностью построенной в интерфейсе браузера. Это же относится к созданию полнофункционального виртуального музея.

Литература

- [1] Анна Турчик. Мультимодальное взаимодействие: исследовательские возможности применения конверсационного анализа // СОЦИОЛОГИЧЕСКОЕ ОБОЗРЕНИЕ. Т. 10. № 1-2. 2011 URL: <https://www.hse.ru/data/2012/10/31/1246279796/text1.pdf> (дата обращения: 15.04.2016).
- [2] Сервис-ориентированная архитектура // URL: https://ru.wikipedia.org/wiki/Сервис-ориентированная_архитектура (дата обращения: 15.04.2016).
- [3] Google Earth (обзор) // URL: http://itc.ua/articles/google_earth_22033/ (дата обращения: 15.04.2016).
- [4] Ляпин С.Х. Сервисы электронной полнотекстовой библиотеки для образования, науки и культуры // Научная периодика: проблемы и решения — 2(14), 2013 март-апрель. URL: <http://www.dilib.ru/journal/articles/50865.php>; Дата обращения 23.03.2015.
- [5] Ляпин С.Х., Куковьякин А.В., Мбого И.А., Толстикова И.И., Чугунов А.В. СЕРВИСЫ ПОЛНОТЕКСТОВОГО ПОИСКА В РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СРЕДЕ (ПРОЕКТ HUMANITARIANA) // НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (Санкт-Петербург), Т.15, №1, 2015г. С. 155-162. DOI: 10.17586/2226-1494-2015-15-1-155-162; URL: <http://elibrary.ru/item.asp?id=23028758> (дата обращения: 26.01.2016.) На сайте Университета ИТМО: <http://ntv.ifmo.ru/file/article/11451.pdf> (дата обращения: 26.01.2016.)
- [6] Metasearch engine // [Эл. ресурс]: http://en.wikipedia.org/wiki/Metasearch_engine. (Дата обращения 16.04.2016)
- [7] Э. Таненбаум, М. ван Стеен. Распределенные системы. Принципы и парадигмы. – СПб.: Питер, 2003. Пер. с английского В. Горбунков. – 877 с. (с. 23).

Lomonosov's Universum: the experience for functional integration of virtual Atlas and full-text library

S. Kh. Lyapin¹, A.V. Kukovyakin²

¹ITMO University, ²Constanta, Ltd

Abstract. Describes the experience of creating a functionally integrated distributed multimodal information medium comprising: a) an electronic library with full-text search resources, related to the activities of M. V. Lomonosov; b) virtual Atlas of the geographical locations in Russia and abroad, associated with the name of M. V. Lomonosov, c) functional interaction of services of a virtual Atlas with the services of electronic full-text library. The virtual Atlas is created based on the technology of Google Earth; virtual globe identified the places associated with the life and work of M. V. Lomonosov (about 100 locations), are created thematic layers, active objects which interact with the thematic search on full-text electronic library. The report was prepared with the support of a grant from the Russian humanitarian scientific Fund No. 14-03-12017

Keywords: full-text search; multimodal interaction; service-oriented architecture; virtual hypermedia Atlas; functional integration of services; paragraph-based query.

References

- [1] Anna Turchik. Mul'timodal'noe vzaimodejstvie: issledovatel'skie vozmozhnosti primeneniya konversacionnogo analiza // Sociologicheskoe obozrenie. T. 10. № 1-2. 2011 URL: <https://www.hse.ru/data/2012/10/31/1246279796/text1.pdf> (data obrashhenija: 15.04.2016).
- [2] Servis-orientirovannaja arhitektura // URL: https://ru.wikipedia.org/wiki/Servis-orientirovannaja_arhitektura (data obrashhenija: 15.04.2016).
- [3] Google Earth (obzor) // URL: http://itc.ua/articles/google_earth_22033/ (data obrashhenija: 15.04.2016).
- [4] Ljapin S.H. Servisy jelektronnoj polnotekstovoj biblioteki dlja obrazovanija, nauki i kul'tury // Nauchnaja periodika: problemy i reshenija — 2(14), 2013 mart-aprel'. URL: <http://www.dilib.ru/journal/articles/50865.php>; Data obrashhenija 23.03.2015.
- [5] Ljapin S.H., Kukovjakin A.V., Mbogo I.A., Tolstikova I.I., Chugunov A.V. Servisy polnotekstovogo poiska v raspredelennoj informacionnoj srede (Proekt HUMANITARIANA) // Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki, Sankt-Peterburgskij nacional'nyj issledovatel'skij universitet informacionnyh tehnologij, mehaniki i optiki (Sankt-Peterburg), T.15, №1, 2015g. S. 155-162. DOI: 10.17586/2226-1494-2015-15-1-155-162; URL: <http://elibrary.ru/item.asp?id=23028758> (data obrashhenija: 26.01.2016.) Na sajte Universiteta ITMO: <http://ntv.ifmo.ru/file/article/11451.pdf> (data obrashhenija: 26.01.2016.)
- [6] Metasearch engine // [Jel. resurs]: http://en.wikipedia.org/wiki/Metasearch_engine. (Data obrashhenija 16.04.2016)
- [7] Je. Tanenbaum, M. van Steen. Raspredelennye sistemy. Principy i paradigmy. – SPb.: Piter, 2003. Per. s anglijskogo V. Gorbunkov. – 877 s. (с. 23).