

Концепт-ориентированный поиск в электронной полнотекстовой библиотеке с мультимодальным расширением

© С.Х.Ляпин, А.В.Куковякин

НП Центроконцепт, Архангельск, Россия
lyapin@atknet.ru

Аннотация

Описывается мультимодальное (соответственно – функциональное) расширение электронной библиотеки T-Libra, обладающей возможностями гибкого параметризуемого полнотекстового поиска и реализованной в архитектуре Web-browser / Web-server / SQL-server. Методологической основой расширения является теория смысловых единиц информации (СЕИ), или мультимодальных концептов (ММ-концептов). Содержательной и технологической основой функциональной интеграции получающейся мультимодальной среды (в которой могут быть представлены как текст, так и нетекстовые модальности информации) выступает *концепт-ориентированная поисковая система*. Она включает в себя: 1) специальную организацию текстовых метаданных; 2) статическую и/или динамическую «библиотеку концептов»; 3) механизм нелинейного каскадного запроса, соединяющего результаты промежуточного полнотекстового поиска, релевантные тезаурусы, библиотеки концептов, а также нетекстовые объекты различной модальности (графика, звук, видео и т.д.); 4) специализированную подсистему презентации результатов поиска в среде Веб-браузера. Этот подход предполагается использовать для создания унифицированной информационно-поисковой среды, сочетающей в себе функции библиотеки, архива, музея, образовательного сервера и исследовательской лаборатории.

1. Постановка вопроса: от электронной полнотекстовой библиотеки к мультимодальной информационной среде

Электронная библиотека T-Libra в архитектуре Web-browser / Web-server / SQL-server с возможностями гибкого параметризуемого поиска по полнотекстовой базе данных, при поддержке пополняемых электронных словарей и под управлением СУБД Sybase ASA v.7.0. (дополненной специально разработанной объектно-ориентированной инструментальной средой X-Taurus) создана в ООО «Константа» и НП «Центроконцепт» в 2000-2003 гг.

Прототип библиотеки был представлен на RCDL'02 в Дубне, полнофункциональная действующая версия 5.0. – на конференции «Культура: от информации к знанию» в Москве в апреле 2003; версия 5.1. – на конференции «Научный сервис в сети Интернет», Абрау-Дюрсо, сентябрь 2003; версия 5.2. – на RCDL'03 [1], варианты бимодального расширения библиотеки – на международной конференции EVA'2003 [2] и конференции АДИТ в Самаре, июнь 2004 [3].

Версии 5.0. и 5.1. используются в Институте управления (г. Архангельск) и в его филиалах в рамках как аудиторной, так и дистанционной технологий обучения. В настоящее время на платформе T-Libra ведется разработка нескольких специализированных региональных библиотек (для медицины, экологии, культуры), а также образовательного сервера для поддержки дополнительного и профессионального образования.

При этом архитектура создаваемых библиотек проектируется с использованием различных вариантов СУБД (Sybase ASA, MS SQL Server, Oracle SQL Server, MySQL) и двух вариантов бизнес-логики приложений (внутренней логики, существенно использующей особенности этих СУБД и реализуемой на языке SQL, и внешней логики, встроенной в CGI-расширение Веб-сервера

и реализуемой на объектно-ориентированном языке C++).

Ход этих разработок показывает, что методология и технология, реализованные при создании линии T-Libra, могут быть использованы в качестве платформы для мультимодального (соответственно – функционального) расширения и поэтапного создания гибкой интеллектуальной информационной среды, сочетающей в себе функции библиотеки, архива, виртуального музея, образовательного сервера и исследовательской Интернет-лаборатории.

Однако для этого недостаточно простого расширения ресурсной базы библиотеки, то есть добавления к ее полнотекстовым файловым и реляционным базам данных нетекстовых модальностей информации (графика, звук, видео и т.д.).

Необходимо решить, концептуально и технологически, на основе какой информационной модальности (или модальностей) будет происходить интеграция всех ресурсов, а также как будет выглядеть унифицированная система поиска и презентации его результатов в мультимодальной информационной среде.

В докладе обосновываются следующие положения:

а) в качестве основы для интеграции ресурсов различной информационной модальности может быть избран *текст*, то есть (*полно*)*текстовая модальность информации*, а в качестве основы для создания унифицированной поисковой системы в мультимодальной среде – гибкая параметризуемая система поиска по текстовым метаданным и полнотекстовым ресурсам, существующим в формате SQL-баз данных;

б) для реализации этого подхода «текст» должен рассматриваться с широкой семиотической точки зрения, как универсальная самоорганизующаяся среда обитания и генерирования разнообразных культурных смыслов [4], как управляемый ресурс для экспликации, трансформации и генерации знания [5];

в) для этого, в свою очередь, различные наблюдаемые и технологически доступные единицы текста (отдельные слова и термины, линейные словосочетания и нелинейные терминологические кластеры, предложения, авторские абзацы и т.д.) следует рассматривать в качестве одно- или многократно редуцированных форм существования более общих многомерных и мультимодальных смысловых единиц информации, – *мультимодальных концептов (ММ-концептов)*;

г) при вышеуказанных предпосылках гибкая функциональная интеграция получающейся мультимодальной информационной среды может быть обеспечена *концепт-ориентированной поисковой системой*, которая включает в себя: 1) специальную организацию текстовых метаданных; 2) статическую и/или динамическую «библиотеку концептов»; 3) механизм нелинейного каскадного

запроса, соединяющего результаты промежуточного полнотекстового поиска, релевантные тезаурусы, библиотеки концептов, а также нетекстовые объекты различной модальности (графика, звук, видео и т.д.); 4) специализированную подсистему презентации результатов поиска в среде Веб-браузера.

2. Многомерные мультимодальные концепты (ММ-концепты) как смысловые единицы информации.

Термин как редуцированный концепт

Многомерными мультимодальными концептами (ММ-концептами) в рамках развиваемой нами с начала 1990-х годов общей теории концептов, или концептологии [6] мы называем *смысловые единицы информации (СЕИ)*, существующие в широком культурно-историческом пространстве, но опирающиеся на понятийный (или псевдо-, или пред-понятийный) базис, закрепленный в значении какого-либо наблюдаемого культурного знака: научного термина, или слова (словосочетания) обыденного языка, или более сложной семантической структуры (например, нелинейного терминологического кластера), или невербального предметного (квазипредметного) образа, или предметного (квазипредметного) действия и т.д.

Мы употребляем здесь выражение «ММ-концепт», чтобы одновременно подчеркнуть и близость этого термина к общеупотребительному термину «концепт», и его отличие, связанное с принципиальной многомерностью и, соответственно, информационной мультимодальностью обозначаемой этим выражением смысловой единицы информации. В современном общенаучном языке термин «концепт» (*das Konzept* - нем., *concept* - англ.) практически всегда по смыслу очень близок к «понятию» или к некоторому «конспекту» (в немецком языке), свернутому «описанию», «концепции», «теории» [7]; в современной лингвистике «концепт» языкового явления чаще всего есть просто логическая пропозиция этого явления.

Ближе всего по методологическому замыслу наше понимание ММ-концепта к тому обобщенному значению, которое имеет термин «*conceptus*» у Петра Абеляра и Фомы Аквинского, вообще в средневековом концептуализме [8]. Прямое отношение к развиваемому нами подходу (в том числе и в понимании «текста» как самоорганизующегося культурного феномена) имеет направление французского постмодернизма, прежде всего связанное с именем Ж.Делёза [9], а также фундаментальное исследование акад. Ю.С.Степанова, в котором концепты трактуются как культурные константы [10].

Главное в ММ-концепте, как мы его понимаем, – это дискретная целостность многомерного и мультимодального смысла, существующего, однако, в непрерывном культурно-историческом

пространстве и поэтому предрасположенного к культурной (и культурогенной!) трансляции из одной предметной области в другую; эти предметные области могут находиться, с дисциплинарной точки зрения, весьма "далеко" друг от друга (наука и искусство, физика и лингвистика, этика и математика и т.п.).

Образуется ли многомерная и мультимодальная дискретная целостность смысла, характерная для концепта, за счет взаимодействия «понятия», «образа», «действия» (можно назвать это *классической трехмерной функционально-структурной моделью концепта*), или за счет каких-то других функциональных элементов архитектоники концепта; транслируется ли, далее, эта целостность по модели логического развертывания понятия, эстетического развертывания образа, коммуникативного развертывания действия т.д. – все это вторично (хотя тоже очень важно), и определяется конкретными и ситуативными обстоятельствами человеческой деятельности.

С этой точки зрения «термин» (ключевая информационная единица для системы полнотекстового поиска) или терминологический кластер представляют собой *одно- или многократно редуцированный культурный концепт* [11]. В рамках развитаемого нами подхода необходимо найти приемлемое технологическое решение для восстановления (или генерации) ММ-концепта из его терминологически редуцированной текстовой формы, а также релевантных нетекстовых модальностей информации (графика, звук, видео и т.д.), имеющих терминологически выраженное метаописание.

Примечание. Мы не касаемся в данной публикации сложной проблемы прямого распознавания и смысловой идентификации "образа" ("действия" и т.п.) и восстановления концепта на этой основе. Описание одного из подходов к постановке этой проблемы, связанного с так называемыми "образными знаками" для индексирования образной (географической, картографической) информации, см. в [12].

3. Гибкий параметризуемый полнотекстовый поиск как основа для построения концепт-ориентированных запросов

Специализированная подсистема «Полнотекстовый поиск» ЭБ Т-Libra организована как совокупность параметризуемых запросов к полнотекстовой SQL-базе данных. Подсистема ориентирована на компьютерную поддержку тех видов деятельности пользователя, в основе которой лежит гибкая тематическая обработка полнотекстовых ресурсов (статей, монографий, учебников и т.п.).

Все реализованные в Т-Libra запросы предназначены для экспликации *терминологических полей* разного типа, вида и уровня.

«Терминологическими полями» мы называем фрагменты смыслового пространства, которые соотношены в самом тексте или в результатах его обработки с релевантными терминологическими структурами: линейно упорядоченной совокупностью терминов в рамках предложения, авторского абзаца или более обширного авторского дискурса; многослойной терминологической структурой, каждый из «слоев» которой представляет некоторый аспект темы или проблемы; частотно-ранжированными списками терминов («терминограммами») отдельного произведения или их совокупности и т.д.

Термины выступают здесь в роли своеобразных точечных смысловых «зарядов», а их различным образом упорядоченные совокупности репрезентируют сложноорганизованные смысловые поля, в явном или чаще неявном виде содержащиеся в первичном тексте.

В архитектуре подсистемы предусмотрены: а) блок электронных словарей с программой автоматической генерации лексико-грамматических парадигм естественного языка (в версии 5.2. имеется около 1 млн. словоформ русского языка); б) блок гибких параметризуемых поисковых запросов (в версии 5.2. представлены 2 типа и 6 видов таких запросов); в) блок функциональных моделей смыслового (концепт-ориентированного) поиска (в версии 5.2. не представлен).

Методологически и технически в системе запросов выделяются: а) *единицы поиска* (в зависимости от вида запроса это могут быть термины, цепочки терминов, многослойные терминологические структуры и т.д.), б) *единицы результатов поиска* (произвольный абзац, частотно-ранжированная совокупность терминов и т.д.), в) *единицы представления результатов поиска* (авторский абзац, таблица частотно-ранжированных терминов и т.д.).

Для каждого из запросов имеется возможность формировать «корзину ресурсов», то есть область определения запроса, в пределах которой он выполняется.

В версии 5.2. реализованы 2 типа запросов: «Абзацно-ориентированный поиск» и «Частотно-ориентированный поиск», каждый из которых представлен тремя видами. Первый тип запросов ориентирован на экспликацию линейно упорядоченных терминологических полей в пределах естественной смысловой единицы текста – авторского абзаца. Второй тип – на частотный анализ текста и построение частотно-ранжированных таблиц терминов, «терминограмм» произведения или совокупности произведений.

Запросы типа «Абзацно-ориентированный поиск»

Простой (однослойный) терминологический поиск. Производится по одному или нескольким

терминам, образующих тем самым некоторый условный тематический «слой». Результатом запроса является совокупность абзацев (включая абзацы сносок и примечаний) из выбранного пользователем произведения, содержащих указанные термины. В найденных абзацах указанные пользователем термины выделены цветом. Поиск ведется с учетом лексико-грамматических признаков слова. Это относится также ко всем типам и видам запросов, в которых простой (однослойный) терминологический поиск входит в их состав.

Кластерный терминологический поиск. Пользователь на выбранной им предметной области <авторы, тексты> задает: а) список терминов, репрезентирующих некоторую тему (проблему, аспект и т.п.), и б) минимально необходимое количество терминов из этого списка, учитываемое при запросе. Результатом запроса является совокупность абзацев, содержащих как минимум заданное количество терминов («терминологический кластер»). Запрос позволяет эксплицировать тему (проблему, аспект и т.п.), неявно определяемую исходным списком терминов, и варьировать ее актуализацию путем изменения как содержания и длины самого списка, так и размеров терминологического кластера, учитываемого при поиске.

Расширенный (многослойный) терминологический поиск. Производится по нескольким терминам, которые одновременно: а) принадлежат разным тематическим «слоям» терминов, при этом количество слоев варьируется и б) находятся на определенном «расстоянии» между собой. Результатом запроса является совокупность авторских абзацев из выбранного произведения, содержащих указанную многослойную терминологическую структуру.

Запрос позволяет на некотором терминологическом поле, определяемом всеми терминами запроса, фиксировать смысловую связь между терминами, относящимися к разным слоям. Изменяя пять параметров: 1) содержание списков терминов в каждом из слоев, 2) количество терминов, включаемых в каждый из слоев, 3) количество слоев, включаемых в запрос, 4) количество слоев, реально учитываемых при поиске, 5) расстояние между терминами, принадлежащими разным слоям, можно производить смысловую фокусировку поиска, варьировать смысловую плотность эксплицированного терминологического поля.

Запросы типа «Частотно-ориентированный поиск»

Абсолютный частотный поиск. Создает частотно-ранжированный список слов (существительных) по «корзине ресурсов» (произведению или нескольким произведениям), собранной из первичных ресурсов. При поиске учитываются все русскоязычные словоформы, которые затем приводятся к нормальной форме:

существительное в именительном падеже, единственном числе. Результатом запроса является таблица, содержащая столбец с частотно-ранжированным списком слов, а также столбец с указанием абсолютной и относительной частоты встречаемости слова (в промилле). Глубина ранжирования может варьироваться пользователем (в диапазоне от 1 до 200 слов); по умолчанию она равна 30 словам.

Запрос дает своеобразную «терминограмму» произведения или их совокупности, неявно определяющую их общую предметно-тематическую ориентацию. Он позволяет текстологически сравнивать друг с другом несколько произведений и/или авторов (путем сравнения соответствующих списков слов), а также сравнивать объективно полученные терминограммы и субъективные экспертные списки терминов.

Авторские терминограммы. Является частным случаем предыдущего вида запроса и создает частотно-ранжированные списки слов всех произведений указанного автора, содержащихся в полнотекстовой БД. Запрос косвенно определяет предметно-тематическую направленность исследований автора, а также позволяет текстологически сравнивать друг с другом различных авторов, или сравнивать авторские терминограммы и релевантные экспертные списки терминов.

Относительный частотный поиск. Для указанного пользователем: а) опорного термина и б) произведения или совокупности произведений запрос сначала выбирает те абзацы, которые содержат этот термин, а затем производит на заданную глубину частотное ранжирование существительных, содержащихся в выбранном подмножестве абзацев.

Запрос позволяет эксплицировать терминологические поля, соотнесенные с термином, выбранном в качестве опорного (последний в итоговой таблице выделен полужирным шрифтом). В данном виде запроса реализуется методология «относительной системы отсчета» по сравнению с методологией «абсолютной системы отсчета», характерной для двух предыдущих видов запросов.

4. Концепт-ориентированный запрос как гибкий функциональный интегратор мультимодальной информационной среды (общая схема)

Исходной точкой построения концепт-ориентированного запроса является один из запросов к полнотекстовой SQL-базе данных, позволяющий по заданному термину или терминологическому кластеру эксплицировать терминологические поля и связанный с ними смысловой микроконтекст в пределах произвольного авторского абзаца (как это описано выше) или отдельного предложения (реализовано в одной из исследовательских версий).

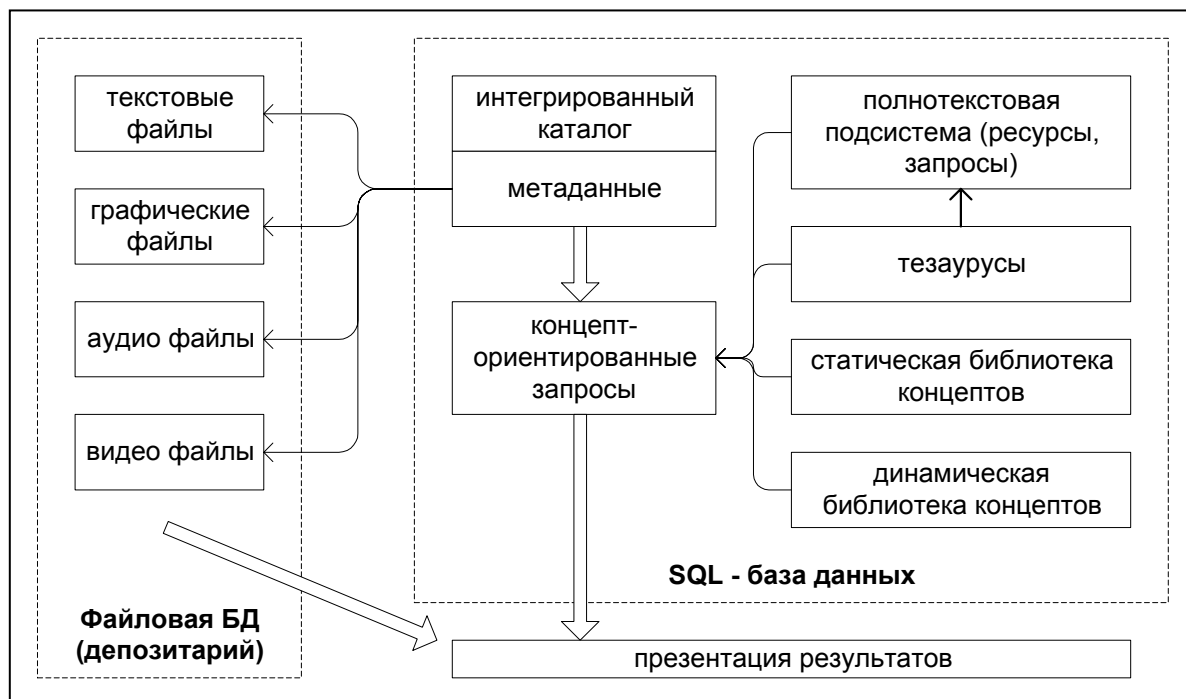


Рисунок 1: Блок-схема ЭБ Т-Libra с концепт-ориентированным запросом и мультимодальным расширением

Результат этого промежуточного запроса – та или иная терминологическая структура, своеобразный терминологический инвариант содержания некоторого множества авторских абзацев, – становится, в свою очередь, началом нелинейного каскадного запроса, выполняющегося в мультимодальной информационной среде с активным использованием текстовых метаданных (их организация или, возможно, динамическая реорганизация в ходе выполнения запроса – предмет отдельного исследования).

В ходе выполнения каскадного запроса для предварительного моделирования предметной области применяется частотный анализ текста по выбранной корзине ресурсов, актуализируются релевантные мини-тезаурусы, а также статическая и/или динамическая «библиотеки концептов».

Заметим, что сами эти тезаурусы различного уровня, как и статические «библиотеки концептов» также могут заранее создаваться с помощью многослойного параметризуемого запроса к полнотекстовой SQL-базе данных (см. описание в п.3), и уже затем в готовом виде использоваться при выполнении собственно концепт-ориентированного запроса.

Что касается динамической «библиотеки концептов», то программные агенты, определяющие ее специфику, запускаются по ходу выполнения запроса и зависят от выбора предметной области и характера промежуточных результатов запроса.

Можно сказать, что в целом в рамках данного подхода используется некоторое обобщенное онтологическое моделирование предметной области, в которой традиционное построение

тезауруса [13] дополняет ту или иную специализированную онтологию [14], или же релевантный тезаурус просто включается в качестве функциональной части в гибридную онтологическую модель, как она понимается, например, в [15].

ММ-концепт выступает здесь в качестве многомерной онтологической «точки» для локальной актуализации и функциональной интеграции мультимодальной информационной среды.

Завершается каскадный концепт-ориентированный запрос презентацией его результата. Соответствующая подсистема обеспечивает отображение в среде Веб-браузера: а) совокупности релевантных текстовых единиц информации (предложений, абзацев), б) совокупности релевантных нетекстовых единиц информации (графических образов и т.п.). в) схематизированного описания полученного мультимодального концепта, которое должно быть адаптировано к функциональному назначению информационной среды (исследование, обучение, музейная экспозиция и т.д.).

5. Архитектура ЭБ Т-Libra с мультимодальным расширением и концепт-ориентированным поиском

Таким образом, основным инструментом мультимодального (соответственно - функционального) расширения электронной полнотекстовой библиотеки является *концепт-ориентированная информационная система*,

соединяющая «текст» и нетекстовые модальности информации как в содержательном плане (через библиотеки концептов и релевантные тезаурусы), так и в технологическом плане (через механизм нелинейного каскадного концепт-ориентированного запроса и презентации его результата).

В общем случае эта система включает в себя:

- а) специальную организацию метаданных для описания первичных ресурсов как текстовой, так и нетекстовых модальностей;
- б) специализированные поля интегрированного каталога ЭБ Т-Libra;
- в) статическую «библиотеку концептов», т.е. набор готовых терминологических и квази-терминологических структур, представляющих собой функциональные модели концептов и предназначенных для включения их в запрос;
- г) динамическую «библиотеку концептов», т.е. набор программных агентов для формирования новых функциональных моделей концептов в ходе выполнения запроса;
- д) подсистему нелинейных каскадных концепт-ориентированных запросов, соединяющих промежуточные результаты полнотекстового поиска, статическую и (или) динамическую «библиотеку концептов», релевантные тезаурусы и нетекстовые объекты гетерогенной базы данных;
- е) подсистему генерации и презентации мультимодальных культурных феноменов в среде Веб-браузера (см. Рисунок 1).

6. Прототип концепт-ориентированного запроса в однородной полнотекстовой среде (исследовательский вариант)

Для изучения возможностей концепт-ориентированного подхода к построению информационных систем различного назначения представляется разумным вначале использовать *однородную полнотекстовую среду*, не содержащую нетекстовых модальностей информации.

С этой целью был разработан экспериментальный вариант каскадного запроса, сочетающего в себе несколько видов абзацно-ориентированного и частотно-ориентированного поиска.

Ниже приводится краткое описание его функциональной структуры и анализ одного из его результатов.

Исходными данными для этого запроса являются:

- термин, обозначающий искомый концепт, например, "факт", "истина" и т.п. (далее - термин);

- совокупность текстов, используемых для построения терминологической модели искомого концепта, например, тексты русской философии, выбранные из SQL-базы данных (область определения);

- совокупность текстов, по которым будет проводиться экспликация концепта на основе его терминологической модели, например, все тексты Н.О.Лосского (область поиска).

Первый этап: построение терминологической модели концепта.

В первую очередь проводится сужение области определения: из исходных текстов остаются лишь те, в которых термин встречается в заголовках или в списке наиболее частотных существительных (ограничение по минимальной относительной частоте встречаемости слова в произведении задается пользователем).

В суженной таким образом области определения проводится поиск наиболее частотных прилагательных (в пределах предложений, в которых находится термин) и существительных (для них строится два списка: один в пределах предложений, второй - в пределах абзацев; в результат включается их пересечение).

Полученные в результате списки прилагательных и существительных репрезентируют терминологическую модель искомого концепта.

Второй этап: экспликация концепта на основе его терминологической модели.

В области поиска находятся авторские абзацы, в которых присутствует термин или есть хотя бы две пары (существительное + прилагательное) из построенных ранее списков при условии, что расстояние между существительным и прилагательным – не более, чем 1 слово.

Затем из найденных абзацев удаляются те, в которых имеются "сторонние" существительные с большей относительной частотой, чем та, которую имеют существительные построенной терминологической модели; пороговое отношение этих частот определяется пользователем.

Совокупность найденных абзацев и представляет собой экспликацию смыслового концепта, обозначаемого исходным термином.

Пример. В качестве опорного термина для экспликации соответствующего концепта и запуска вышеописанного каскадного запроса был выбран "факт", в качестве области определения, на которой стоилась концептуальная модель "факта" – множество русскоязычных философских текстов (около 100 произведений).

Затем полученная модель "фактуальности" была применена к текстам русского философа Н.О.Лосского (книги "Обоснование интуитивизма" и "Мир как органическое целое").

Результатом явилась выборка из примерно 1200 абзацев. При этом сам термин 'факт' со всей его русскоязычной грамматической парадигматикой встречается лишь в 72 из них.

В остальных эксплицируется концептуальный смысл "фактуальности" с разной степенью релевантности (была произведена экспертная оценка результатов запроса): примерно в 280 абзацах (23%) речь идет о фактуальности или прямых концептуальных синонимах 'факта' ('эмпирические данные', 'опытные данные', 'результаты наблюдения' и т.п.), еще примерно в 300 абзацах (25%) эксплицирован ближний смысловый

контекст фактуальности, еще примерно в 400 (33%) – дальний смысловой контекст. Не имеющими смысловой связи с "фактуальностью" были признаны около 220 абзацев (19%).

Разумеется, речь идет лишь о начальном этапе движения к конечной цели: экспликации (из текстов, как в приведенном примере, из текстов + нетекстовых модальностей информации, как следует из общей постановки задачи в п.4) концептов разного типа, вида и уровня, а также об использовании механизмов этой экспликации для последующей функциональной настройки информационной среды.

7. Заключение. От линии T-Libra к семейству T-System

Электронным библиотекам свойственно активно развиваться в направлении многофункциональных информационных сред, с одновременным ростом концептуальности их поисковых и презентационных подсистем; характерно и поучительно в этом плане, например, развитие проекта Alexandria Digital Library [16], [17].

Что касается расширения ЭБ T-Libra, основанного на нелинейных каскадных концепт-ориентированных запросах к мультимодальным базам данных, то мы предполагаем использовать его для построения семейства информационных систем следующего функционального назначения:

- электронных библиотек, сочетающих гибкий полнотекстовый поиск с обработкой и презентацией нетекстовых единиц информации, как содержащихся в самом тексте, так и находящихся за его пределами (собственно линия T-Libra);

- исследовательских систем с возможностями продвинутого поиска по полнотекстовым ресурсам, имеющих в своем составе виртуальные лаборатории, прежде всего гуманитарной направленности (линия T-Research);

- образовательных систем для дистанционного обучения, в которых система концепт-ориентированного поиска и презентации его результатов адаптирована к учебным планам и программам (линия T-Education);

- интерактивных тематизируемых мультимодальных экспозиций в составе, например, виртуальных музеев (линия T-Media).

В целом можно говорить о семействе T-System, использующих «текст» в качестве базисной модальности для функциональной интеграции всей мультимодальной информации, а гибкую концепт-ориентированную поисковую систему, активно взаимодействующую с метаданными – в качестве универсального средства ее обработки.

В настоящее время этот подход реализуется в рамках нескольких пилотных проектов по созданию на базе T-Libra специализированных информационных систем (виртуальной Интернет-лаборатории гуманитарной направленности; образовательного сервера для сфер общего,

профессионального и дополнительного образования; виртуального музея северной иконы).

Литература

- [1] С.Х.Ляпин, А.В.Куковьякин. Многофункциональная электронная библиотека T-Libra: WWS-архитектура, интегрированный каталог, настраиваемый мультирубрикатор, гибкий параметризуемый полнотекстовый поиск // Труды 5-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL 2003, Санкт-Петербург, Россия, 2003. - Изд-во СПбГУ, 2003, с. 292-299; <http://rcdl2003.spbu.ru/proceedings/J4.pdf>
- [2] С.Х.Ляпин, А.В.Куковьякин. Мультимодальное расширение электронной полнотекстовой библиотеки: концепция, методология, технология // Материалы 6-й ежегодной конференции EVA 2003. Москва, Государственная Третьяковская Галерея, 1-5 декабря 2003 года. http://www.evarussia.ru/upload/doklad/doklad_928.doc
- [3] С.Х.Ляпин, А.В.Куковьякин. T-Media: интегрированная среда для объединения текстовых и нетекстовых информационных ресурсов и создания интерактивных тематических экспозиций // Материалы Восьмой ежегодной конференции АДТИТ-2004, Самара, Самарский областной историко-краеведческий музей им. П.В.Алабина, 31 мая - 5 июня 2004. – Самара, 2004, с. 53-54.
- [4] См.: Иванов Вяч. Вс. Иероглифические, гетерографические и алфавитные способы построения текста и передачи его по времени // Семиотика культуры. Тез. докл. Всес. шк.-семина. по семиотике культуры 8-18 сент. 1988 г. – Отв. ред. Н.М.Теребихин. – Архангельск: АГПИ им. М.В.Ломоносова, 1988. – с.10-13; Лотман Ю.М. К современному понятию текста // Семиотика культуры. Тез. докл. Всес. шк.-семина. по семиотике культуры 8-18 сент. 1988 г. – Отв. ред. Н.М.Теребихин. – Архангельск: АГПИ им. М.В.Ломоносова, 1988. – с.3-6; Лотман Ю.М. О каузальных связях в семиотическом ряду // Семиотика культуры. Тез. докл. Всес. шк.-семина. по семиотике культуры 8-18 сент. 1988 г. – Отв. ред. Н.М.Теребихин. – Архангельск: АГПИ им. М.В.Ломоносова, 1988. – с.6-10; Текст: семантика и структура. – М.: Наука, 1983. – 302с; Текст: структура и анализ. – М.: Изд-во ИЯ АН СССР, 1989; Топоров В.Н. Пространство и текст // Текст: семантика и структура. – М.: Наука, 1983.– 302с. (с.227-285).
- [5] Ляпин С.Х., Скидан О.П., Куковьякин А.В. Текст как управляемый познавательный ресурс (к экспликации, трансформации и генерации

- знания с помощью концепт-ориентированных информационных систем) // Вестник СЗО РАО. – №3, 1998, СПб.-Архангельск: Изд-во Поморского университета, 1998. – с.195-205.
- [6] См.: *Ляпин С.Х.* О концептах и концептологии (в поисках нового подхода к моделированию деятельности) // XIX World Congress of Philosophy. – Moscow 22-28 August 1993. – Book of abstracts. Сборник резюме. Vol. I. Секция 13 (Философия деятельности). – с.322; *Ляпин С.Х.* Концептология: учение о концептах, методология культурогенных трансляций, технология эвристического развертывания смысла // Вестник СЗО РАО. – №3, 1998, СПб.-Архангельск: Поморский гос. университет им. М.В.Ломоносова, 1998. – с.28-41. *Ляпин С.Х.* Концептологическая формула факта // Концепты. Научные труды Центроконцепта. – Отв. ред. С.Х.Ляпин. – Вып. 2(2) 1997. – Архангельск: Изд-во Поморского гос. университета, 1997. – с.5-71.
- [7] См., напр.: поиск по термину 'Concept formation' // Encyclopaedia Britannica CD 2.0, 1995; *Maria Pozzi.* The Concept of 'Concept' in Terminology: a Need for a New Approach. In: Proc. of the 5th Int. Congress on Terminology and Knowledge Engineering, 23-27 August, 1999, Innsbruck, Austria, pp. 28-42; *Van Mechelen, J., Hampton, J.A., Michalski, R., & Theuns, P.* (Eds.), Categories and Concepts: Theoretical Views and Inductive Data Analysis, Academic Press, London, UK.
- [8] См.: *С.С.Неретина.* Слово и текст в средневековой культуре. Концептуализм Абеяра. - М., Гнозис, 1993; *С.С.Неретина.* Верующий разум. К истории средневековой философии. - Архангельск: Изд-во Поморского межд. пед. университета, 1995.
- [9] *Жиль Делёз, Феликс Гваттари.* Что такое философия? Пер. с фр. С.Н.Зенкина. – М., «Институт экспериментальной социологии»; СПб., Изд-во АЛТЕЙЯ, 1998.
- [10] *Ю.С.Степанов.* Константы: Словарь русской культуры. – Изд. 2-е, испр. и доп. – М.: Академический проект, 2001.
- [11] *Sergei Kh. Lyapin.* Term as reduced concept (to conceptology analysis of science terminology) // Proc. of 5th Int. Congress on Terminology and Knowledge Engineering, 23-27 August 1999, Innsbruck, Austria. pp. 89-101.
- [12] *И.М.Защман.* Визуально-мотивированное представление знаний в электронных библиотеках научных документов // Труды 4-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL 2002, Дубна, Россия, 2003, с.120-135.
- [13] *Аджиев Алим Сапарович, Нгуен Хунг Мань.* Подходы к описанию и использованию тезаурусов в информационных системах // Труды 5-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL2003, Санкт-Петербург, Россия, 2003, с.191-200.
- [14] *Pavel Smrz, Anna Sinopalnikova, Martin Povolny.* Thesauri and Ontologies for Digital Libraries // Proceedings of the 5th Russian Conference on Digital Libraries RCDL2003, St.-Petersburg, Russia, 2003, p.14-16.
- [15] *Leonid Kalinichenko, Michele Missikoff, Federica Schiappelli, Nikolay Skvortsov.* Ontological Modeling // Proceedings of the 5th Russian Conference on Digital Libraries RCDL2003, St.-Petersburg, Russia, 2003, p.7-13.
- [16] *Terry Smith, Alex Ushakov, Bill Heller.* Some Aspects of Developing and Using the Digital Learning Environment in Alexandria Digital Earth Prototype // Proceedings of the 5th Russian Conference on Digital Libraries RCDL2003, St.-Petersburg, Russia, 2003, p.18-25.
- [17] *O.Agapova, R.Mayer, T.Smith, A.S.Ushakov, A.A.Ushakov, Stefan Decker.* Developing Digital Library Visual Services for Building a Lesson-Design Environment Prototype // Proceedings of the 5th Russian Conference on Digital Libraries RCDL2003, St.-Petersburg, Russia, 2003, p.130-139.

Concept-Oriented Search In Digital Full-Text Library With Multi-Modal Extension

Sergey Lyapin, Alexey Kukovyakin

We describe hereby multi-modal (and related multi-functional) extension of the digital library T-Libra, which has possibilities of flexible parameterized full-text search and that is realized in Web-browser / Web-server / SQL-server architecture. The methodological basis for extension is the theory of sense units of information (SUI), or multi-modal concepts (MM-concepts). *Concept-oriented search system* here is the substantial and technological basis for functional integration of resulting multi-modal environment (where both text and non-text modalities of information can be presented). This search system includes the following: 1) special organization of text metadata, 2) static and/or dynamic "library of concepts", 3) mechanisms of nonlinear cascade inquiries which combine the results of intermediate full-text search, relevant thesauri, libraries of concepts as well as non-text objects of different modalities (graphics, sound, video etc.), 4) specialized subsystem for presentation of search results in Web-browser's environment. We propose to use this approach for developing of unified information-searching environment that combine functions of library, archive, museum, education server and research laboratory.