

Технологии электронных библиотек для поддержки справочных систем в медицине

© Д.Э. Палей, Д.Н. Курчинский, Ю.В. Широков, В.Н. Смирнов

Ярославский государственный университет им. П.Г. Демидова

paley@yars.free.net, reno@econom.uni Yar.ac.ru, shirokov@jugra.yar.ru, smirnov@uniyar.ac.ru

Аннотация

Работа посвящена описанию использования технологий, разработанных и применяющихся при создании цифровых каталогов и электронных библиотек для построения разнообразных систем. В качестве примера такой системы рассмотрен программный комплекс «Аптечная справка», предназначенный для сбора и поиска информации о наличии и ценах на лекарственные средства в г. Ярославле.

1 Введение

Исторически так сложилось, что авторы доклада уже много лет в разных качествах участвуют в разработке и реализации нескольких информационных систем. Поскольку время работы над этими проектами уже измеряется годами (что, по нашему мнению, достаточно долго), возникла потребность проанализировать, какие же технологии, приемы моделирования и разработки показали наибольшую эффективность и применяются наиболее широко. Следует сказать, что проекты достаточно разноплановые по предметным областям. В качестве основных можно указать такие, как электронные библиотеки (ЭБ) [1, 2], автоматизация классических библиотек [3], информатизация музеев [4], медицинские информационные системы [5], CRM-системы для рекламных компаний и т. д.

И вот, даже при беглом взгляде, выяснилось, что технологическую основу большинства разработок составляют методы и приемы, пришедшие из электронных библиотек. Факт этот для нас как очевиден, так и неожиданно удивителен, т. к. проекты развивались абсолютно независимо и прошли за время своего существования через ряд серьезных переработок.

В этой работе представлен один из наших проектов, и на его примере показаны возможности ис-

пользования технологий ЭБ в практике разработки и внедрения информационных систем. Описываемый проект называется «Аптечная справка». Основной его задачей является сбор данных о ценах и наличии лекарственных средств в аптеках города Ярославля, предоставление информации населению, а также формирование различной аналитической отчетности по собранным данным.

Обратимся к «истории вопроса».

2 Немного истории

2.1 Цифровые библиотеки

Как уже было сказано выше, авторы доклада достаточно много лет занимаются разработкой и внедрением электронных библиотек различного типа в учебных и научных учреждениях. В сфере интересов в основном находились информационные системы, в задачи которых входила обработка информации об объектах, описательная структура которых не четко определена или может меняться с течением времени. Это предполагало решение дополнительных задач по обработке метаданных таких электронных библиотек. Технологии, применяемые в работах, предполагали, что описание моделируемой предметной области осуществляется на основе того же цифрового каталога, в котором хранятся данные самих объектов [1].

2.2 Библиотечные и музейные системы

Также выполнен ряд работ по разработке информационных систем для библиотек и музеев [4, 6–8]. Применяемые здесь технологии, имеют много особенностей, связанных с предметной областью. Вместе с тем они достаточно много заимствовали из проектов ЭБ. Более того, в последнее время тесная интеграция этих систем стала одним из основных направлений деятельности [9].

Суммируя эти сведения, можно сказать, что выполнено много работ, накоплены определенные методические приемы, опыт, умения и знания.

3 Проект «Аптечная справка»

Основная цель проекта – сбор данных о наличии и ценах на лекарственные средства (ЛС) в аптеках города, предоставление их населению, анализ полученных данных по различным критериям. Старт проекта состоялся в 1995 году, и с тех пор он существует и развивается. [5]

Понятно, что для реализации и поддержки проекта предпринято и осуществляется большое количество организационных и технических мероприятий. В наши цели входит анализ только информационной инфраструктуры, существующей в настоящее время.

Следует сказать, что подобные проекты успешно реализованы и в других регионах России. В качестве примера можно привести информационные системы, существующие в Москве [10, 11] или Екатеринбурге [12]. Каждый такой проект является достаточно специфичным. Очевидно, что его реализация сильно зависит от местных условий, заказчика, возможностей исполнителя и т. д.

Наша система является распределенной. Схема основных потоков информации представлена на рис. 1.

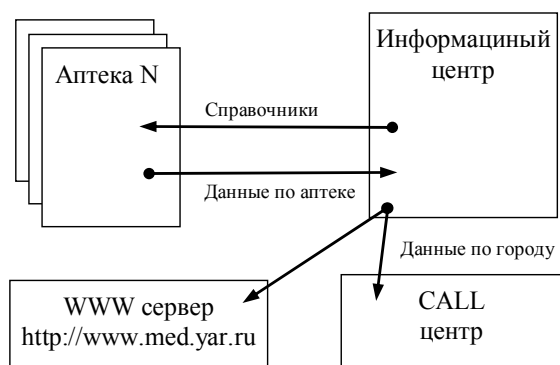


Рис. 1. Информационные потоки в системе

Участники системы (аптеки) передают данные на сервер информационного центра. Это данные по наличию и ценам на ЛС, справочные данные по аптеке (время работы, телефоны, проезд и т. д.), информационные объявления различного характера (реклама, различные акции для покупателей, данные об условиях отпуска медикаментов и т. д.).

С центрального сервера аптеки получают справочные данные, необходимые для работы, аналитическую информацию по предельным ценам, различные нормативные реестры (например, реестр ЛС изъятых из продажи), а также информацию о ценах на ЛС во всем городе.

На основе оперативных данных центрального сервера реализован веб-сервис по поиску ЛС в аптеках города. Этот же сервис позволяет предоставить населению справочную информацию от аптек. Ежедневно оперативные данные добавляются в БД истории.

Параллельно данные сервера поступают в приложение звонкового CALL-центра, где операторы

имеют возможности поиска ЛС, аналогичные веб-приложению. Также система CALL-центра позволяет анализировать оперативные и исторические данные по различным параметрам.

Опишем задачи, которые были решены для обработки каждого из потоков информации, и проблемы, возникшие при их решении. Основной из таких задач является поиск ЛС в данных от аптек.

4 Реализация системы поиска лекарственных средств на основе рубрикаторов

4.1 Основные принципы поиска

Поиск ЛС в базе данных возможен несколькими способами. Самый простейший из них – «просто» поиск по наименованию с предоставлением конечному пользователю списка полученных совпадений. Это очень просто реализуемо и настолько же неэффективно при количестве цен больше нескольких тысяч.

Более эффективные способы связаны с классификацией лекарственных средств. Фактически это приводит к созданию рубрикаторов, к которым тем или иным способом привязаны все ЛС и, соответственно, цены на них. В терминологии ЭБ это означает, что необходимо создать тезаурус – наборы связанных предметных терминов по коллекции. В качестве коллекции выступает множество всех возможных торговых наименований ЛС от разных производителей во всех аптеках. Принципы построения рубрикаторов диктуются практическими потребностями. Основными из них являются следующие: группировка данных по торговым наименованиям без учета производителя и по международным непатентованным наименованиям. Классификация в этом случае представляет собой сопоставление кода ЛС с рубрикой соответствующего тезауруса.

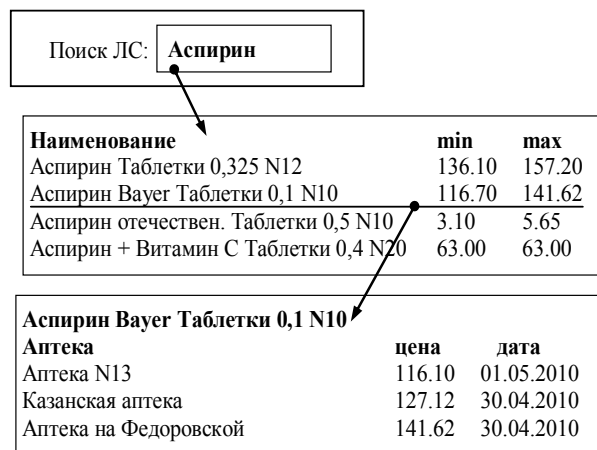


Рис. 2. Структура поиска наличия и цен на ЛС

В настоящее время поиск в системе реализован именно на основе рубрикаторов. Практика показала, что оптимальной является двухуровневая структура, т. е. поисковый запрос первоначально находит подходящие рубрики, и потом пользователь может про-

смотреть конкретные ЛС, входящие в найденную рубрику. Наличие единой классификации ЛС также позволяет еще на этапе начальной обработки данных, полученных от аптек, получить некоторые агрегатные данные по всем ЛС, входящим в рубрику, например, количество различных предложений, минимальную и максимальную цены, начальные и конечные даты получения данных и т. д.

4.2 Задачи построения рубрикаторов

Задача построения рубрикаторов и классификации ЛС осложняется следующими требованиями:

- Список классифицируемых ЛС не фиксируется и может изменяться с течением времени. Новое ЛС может появиться у любого клиента системы (в любой аптеке).

- Каждая аптека должна иметь возможность оперативно предоставить данные по новым ЛС, т. е. привязка к рубрикатору должна быть возможна непосредственно у конечного пользователя.

- Классифицированные в аптеках ЛС должны принадлежать в конечном итоге одним и тем же рубрикам классификаторов.

- Все аптеки объединены в группы (часто это аптечные сети), в которых коды ЛС внутри группы идентичны и, соответственно, их классификация должна совпадать.

- Данные, классифицированные одним из участников системы, должны быть доступны и всем остальным участникам этой же группы.

Учитывая все это, был выбран вариант системы, в котором рубрикатеры формируются централизованно, а классификация производится конечными пользователями (прямо в аптеках). Здесь возникает ряд проблем. Основные из них связаны с выделением ошибок классификации. Опишем эту задачу на простейшем примере.

Положим, у клиента A_n некоторой аптечной сети $A = \{A_i\}$ появилось новое ЛС с кодом L_1 (во всех аптеках сети внутренний код L_1 этого ЛС одинаковый). В аптеке оно было проклассифицировано, т. е. сопоставлено с соответствующими рубриками тезауруса, например, со списком торговых наименований $\{R_j\}$. Данному ЛС **неверно** был сопоставлен некоторый код R_n . Данные об этой классификации поступили на сервер. Теоретически запись сопоставления должна быть предоставлена всем остальным аптекам сети, но допускать этого нельзя, т. к. неизвестно, верна она или нет.

Аналогично положим, что у клиентов A_n, A_m некоторой аптечной сети $A = \{A_i\}$ одновременно появилось новое ЛС с кодом L_2 . В каждой аптеке оно было проклассифицировано. Соответственно, в аптеке A_m данному ЛС был сопоставлен некоторый код R_m , а в аптеке A_n – код R_n . Данные об этой классификации поступили на сервер. В случае, когда коды R_m и R_n совпадают, запись сопоставления должна быть предоставлена всем остальным аптекам сети, хотя сопоставление в обеих аптеках может быть проведено неверно. Если же R_m не равно R_n ,

то возникает неопределенность, которую легко выделить.

Пока одна из записей не «утверждена» или не признана ошибочной, распространять данные по всем аптекам сети нельзя. Очевидно, что принимать решение о правильности классификации может только эксперт – специалист информационного центра. В силу большого количества данных сделать это быстро не представляется возможным.

Вместе с тем обязательным требованием является то, что данные из аптеки **всегда** должны быть доступны для поиска и просмотра, т. к. ошибки классификации могут быть и незначительны, т. е. даже неверным образом классифицированное ЛС должно присутствовать в результатах поиска. На практике это выглядит примерно так: при поиске наименования «аспирин кардио» может быть найден и «аспирин шипучий», и «парацетамол». В этом случае решение об адекватности сопоставления принимается конечным пользователем системы.

4.3 «Временная» классификация и «утверждение» сопоставлений

Описанная выше задача была решена на основе «временной» классификации и статусов «утвержден» для каждого сопоставления ЛС с рубрикатерами. В итоге система автоматической рубрикации выглядит следующим образом.

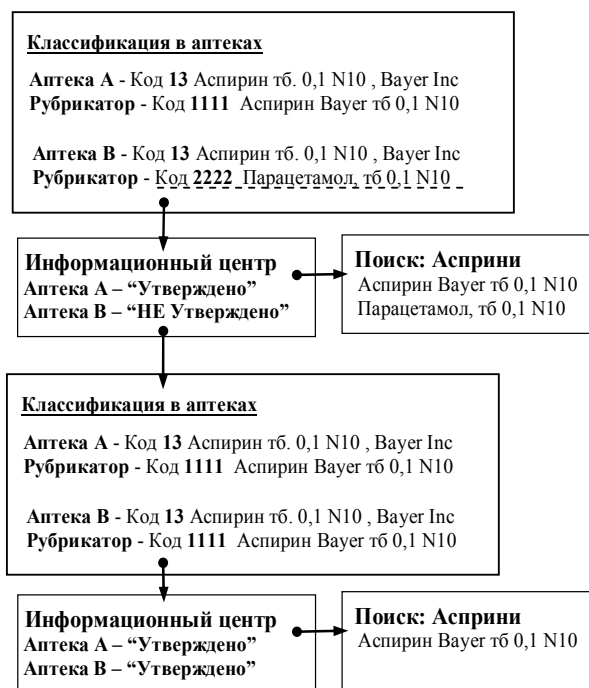


Рис. 3. Схема утверждения классификации аптек одной сети

а) На сервере поддерживаются рубрикатеры и «утвержденные» таблицы сопоставлений с данными от аптек. Эти таблицы поступают в аптеки.

б) Все данные классифицируются непосредственно в аптеках на основе «утвержденных» рубрикатеров и таблиц, а также на основе самостоятельного сопоставления.

с) Данные в аптеки поступают на сервер, где по сопоставлению, отличному от «утвержденного», они записываются в таблицы со статусом «не утвержден».

д) При поиске и представлении данных статус «утвержден» не учитывается, т. е. будет показано все, что соответствует найденной рубрике.

е) Специалисты информационного центра просматривают «неутвержденные» записи сопоставления, если необходимо, исправляют их, присваивают статус «утверждена».

Далее данные поступают в аптеки: утвержденное сопоставление заменяет «неутвержденные» записи.

В заключение приведем несколько количественных характеристик поисковой системы по состоянию на 01.05.2010:

Количество пользователей системы (аптек)	= 107
Количество актуальных цен	~ 250 000
Количество рубрик классификатора ЛС по торговым наименованиям – без производителей	~ 14000
Количество рубрик классификатора ЛС по торговым наименованиям – с производителем	~ 60000
Количество рубрик классификатора ЛС по международным наименованиям	~ 5000

5 Информация от аптек – данные нечеткой структуры

Другой важной целью проекта было создание интернет-каталога аптечных учреждений города, который содержал бы самую разнообразную информацию об аптеках. Это стандартная и не особенно сложная задача. В начале разработки и внедрения системы она и была решена стандартными средствами. Несколько таблиц в базе данных, несложный веб-интерфейс – все это позволило быстро сформировать такой каталог. Но, как это обычно бывает, новые требования породили новые задачи.

Хорошо известно, что лучше всех администрируют данные именно те, кто за них отвечает, т. е. сами аптеки. Была внедрена система сбора данных от аптек.

Прошло время, и стало понятно, что фиксированный формат метаданных для получения информации об аптеке – хорошее решение. Но такое решение, к сожалению, приводит к обновлению версии приложения в каждой аптеке при добавлении любого нового атрибута.

В итоге выяснилось, что наиболее технологично и эффективно решение, применявшееся в цифровой библиотеке. Формально можно представить, что данные об аптеке – это некоторый класс, который описывается набором атрибутов. Для каждого атрибута есть правила ввода и отображения. Клиентское приложение в аптеке получает с сервера информа-

ционного центра описание этого класса и предоставляет работникам аптеки возможности ввода и редактирования всех атрибутов. Аналогичная ситуация с представлением аптеки через веб. Обмен данными производится в XML-формате, что снимает ограничения на модификации структуры в дальнейшем.

Разработка этого направления на момент написания доклада активно развивалась. В планах реализация сервиса по предоставлению клиентам возможностей по формированию собственных информационных атрибутов. В случае успешной реализации каждая аптека или аптечная сеть сможет выставлять на веб-сервере собственные варианты информационных страниц подобно тому, как объекты цифровой библиотеки имеют собственное представление в зависимости от класса, которому они принадлежат [1].

Сейчас часть атрибутов является общесистемными и используется в дополнительных сервисах, например, время работы, список акций для покупателей, наличие тех или иных отделов, расписание и условия приема врачей в аптеках. На веб-сервере просмотр таких атрибутов возможен сразу по всем аптекам. В CALL-центре возможен поиск по этим атрибутам. В дальнейшем предполагается, что общесистемные атрибуты могут быть определены непосредственно в аптеках, например, на уровне аптечных сетей.

6 Классификация данных – дополнительные возможности для анализа

Использование классификаторов позволило организовать различный дополнительный анализ массива цен по разным показателям. Т. к. выбранная рубрикация (по торговым наименованиям ЛС с производителями) фактически наиболее подробная для практических потребностей, то появилась возможность построить дополнительные рубрикаторы на основе существующего.

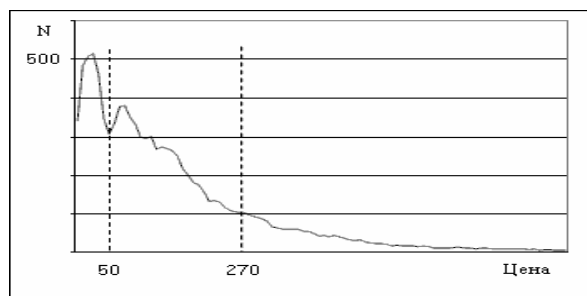


Рис. 4. Распределение предложений противовоспалительных ЛС

Таким образом, дополнительная рубрикация может быть произведена централизованно. В системе «Аптечная справка» существует ряд дополнительных классификаторов: по нозологиям, фармгруппам ЛС и т. д. Соответственно, с их использованием разработана система аналитической отчетности.

Например, на рис. 4 показано распределение предложений в г. Ярославле различных ЛС по диапазонам цен на 01.03.2010. График построен только по одной фармгруппе лекарственных средств (противовоспалительные), исключая биодобавки.

7 Заключение

В данной работе описана прикладная система, основанная на технологии электронных библиотек. Это сервис по поиску лекарственных средств в аптеках г. Ярославля. Основу сбора данных и поисковой машины системы составляет рубрикация, распределенно осуществляемая как в информационном центре, так и конечными пользователями. В работе показаны проблемы, возникающие при таком подходе и пути их решения.

В заключение хочется подчеркнуть, что описанные приемы зарекомендовали свою практическую эффективность и в других медицинских информационных проектах, например: ИС по выписке льготных рецептов, сервис по предоставлению информации о расписании приема врачей в ЛУ и т. д.

Авторы надеются на плодотворное сотрудничество со всеми, кому интересно данное направление развития электронных библиотек.

Литература

- [1] Палей Д.Э., Курчинский Д.Н., Смирнов В.Н. Цифровая библиотека Ярославского региона. Итоги работы, перспективы развития // Труды пятой Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2003, Санкт-Петербург, 29 – 31 октября 2003 г. – Санкт-Петербург: НИИ химии СПбГУ, 2003. – С. 315-319.
- [2] Курчинский Д.Н., Палей Д.Э., Смирнов В.Н. Информационная система учреждения культуры – как система обработки объектов электронного каталога // Труды Седьмой Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2005, Ярославль, 4 – 6 октября 2005 года. – Ярославль: ЯрГУ, 2005. – С. 270-274.
- [3] АБИС «Буки» – официальный сайт. – <http://www.buki.yar.ru>.
- [4] Смирнов В.Н., Смирнова Н.Е., Аносовская А.В. Информационная система по культурному наследию Л.В. Собинова // Труды Всерос. науч. конф. «Научный сервис в сети Интернет», г. Новороссийск, 18 – 23 сентября 2000 г. – М.: Изд-во Московского ун-та, 2000. – С. 150-151.
- [5] Поиск информации по лекарственным средствам. – <http://www.med.yar.ru>.
- [6] Палей Д.Э., Курчинский Д.Н., Смирнов В.Н. АБИС «Буки». Первые итоги развития, новые возможности, перспективы на будущее // Информационные технологии, компьютерные системы и издательская продукция для библиотек: Докл. и тез. докл. – Седьмая Межд. конф. и Выставка «LIVCOM-2003», пансионат «Ершово», Звенигород, Московская область, 17 – 21 ноября 2003 г. – Москва: ГПНТБ России, 2003. – С. 192-200.
- [7] Палей Д.Э., Курчинский Д.Н., Смирнов В.Н. Учет книгообеспеченности в библиотеке ВУЗа: проблемы и пути их решения // Информационные технологии, компьютерные системы и издательская продукция для библиотек: Докл. и тез. докл. – Восьмая Межд. конф. и Выставка «LIVCOM-2004», пансионат «Ершово», Звенигород, Московская область, 15 – 19 ноября 2004 г. – М.: ГПНТБ России, 2004. – С. 155-160.
- [8] Палей Д.Э., Смирнов В.Н., Курчинский Д.Н. Опыт внедрения учета книгообеспеченности в библиотеке ВУЗа // Библиотеки и образование: Сборник материалов первой Межд. конф., Ярославль, 19 – 22 апреля 2005 г. – Ярославль: МУ-БиНТ, 2005. – С. 107-112.
- [9] Курчинский Д.Н., Палей Д.Э., Смирнов В.Н. Электронная библиотека ВУЗа – как инструмент автоматического формирования учебных мультимедийных коллекций // Труды Одиннадцатой Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», RCDL'2009, Петрозаводск, 17 - 21 сентября 2009 года. – Петрозаводск: КарНЦ РАН. – С. 427-432.
- [10] <http://www.poisklekarstv.ru/>.
- [11] <http://www.medlux.ru>.
- [12] <http://www.66.ru/help/pharmacy>.

Digital library technology for medical information systems

D. Paley, D. Kurchinsky, Yu. Shirokov, V. Smirnov

The article describes the medical information system (Drug Information). The main purpose of this system – search drugs in drugstore of Yaroslavl. Digital Library technology is used for objects (drugs) distributed classification. Also distributed classification is main path of system search engine. The positive and negative aspects of this solution and results are discussed in the article.