

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. М. Т. КАЛАШНИКОВА

Информационные технологии и письменное наследие

El'Manuscript-2012

Материалы IV международной научной конференции
Петрозаводск, 3–8 сентября 2012 года

Петрозаводск, Ижевск
2012

УДК 004.9
ББК 81.11+81.2-0
И741

Изданы при поддержке гранта РФФИ (проект № 12-06-06061-г),
гранта РГНФ (проект № 12-04-14154-г) и в рамках реализации комплекса
мероприятий Программы стратегического развития ПетрГУ на 2012-2016 г.

Ответственные редакторы:

В. А. Баранов, д-р филол. наук, проф.

А. Г. Варфоломеев, канд. физ.-мат. наук, доц.

Информационные технологии и письменное наследие [Текст] :
И741 материалы IV междунар. науч. конф. (Петрозаводск, 3–8 сентября
2012 г.) / отв. ред. В. А. Баранов, А. Г. Варфоломеев. — Петрозаводск ;
Ижевск, 2012. — 328 с.

ISBN 978-5-8021-1402-5

Сборник содержит материалы конференции, посвященной современ-
ным электронным средствам хранения, описания, обработки, исследования
и публикации памятников письменности и исторических источников.

УДК 004.9
ББК 81.11+81.2-0

ISBN 978-5-8021-1402-5

© Петрозаводский государственный
университет, 2012
© Ижевский государственный технический
университет им. М. Т. Калашникова, 2012

ПРОГРАММНЫЙ КОМПЛЕКС ПО ОПИСАНИЮ ДОКУМЕНТОВ ЦАРИЦЫНСКОГО ГОРОДОВОГО МАГИСТРАТА КОНЦА XVIII века КАК ИНФОРМАЦИОННАЯ БАЗА РЕКОНСТРУКЦИИ АРХИВНОГО ФОНДА

Н. А. Тупикова, С. А. Преферансов, Д. А. Вуйлов, Г. Л. Цельник

Волгоградский государственный университет, Волгоград

The article is devoted to the description of the integrated program developed on the basis of analysis and classification of the documents, extant as a part of the archival fund of Tsaritsyn police city council (80–90-es of the 18th century). The suggested organization method of the program complex, levels of electronic database, allocated criteria for defining document status and structural-functional properties turn useful parameters for searching the document in frameworks of the archival collection.

Актуальной задачей для прикладной лингвистики считается оптимизация способов фиксации и хранения речевой информации [Кибрик, 2001: 262], что предполагает создание электронных ресурсов, в том числе текстовых массивов [Зубов, 2004: 156]. Документы в составе различных архивных комплексов являются ценными памятниками деловой письменности, изучение которых проливает свет на многие вопросы, связанные с разработкой методологии, методов анализа и эффективного использования документов науки, истории и культуры, определения режима хранения и поиска информации.

«Программный комплекс по описанию документов Царицынского городского магистрата (конец XVIII в.)» позволяет распознавать тексты и классифицировать материал. С точки зрения способа организации данная программа создана на основе файл-серверной архитектуры под названием «Электронный магистрат». Несмотря на существующие недостатки и ограничения данного подхода [Кузнецов, 1998], он позволяет осуществлять многопользовательский режим работы, централизовать управление доступом, имеет низкую стоимость и высокую скорость разработки, а также низкие затраты на поддержание программного продукта. Подобная архитектура рассчитана чаще на небольшие объемы информации и применение в однопользовательском режиме.

Для создания программного продукта важной является дифференциация функциональных компонентов типов информационных систем [Петров, 2002: 30–31], что позволило в программе «Электронный магистрат» представить базу данных и интерфейс приложения. Под интерфейсом приложения следует понимать совокупность электронных форм, логику их обработки и программной логики взаимосвязи с базой данных, включающей операции выборки, вставки, изменения и удаления данных.

Исходя из требований, предъявляемых к приложению в целом (простота развертывания и портативность), в качестве системы управления базами данных (далее — СУБД), удовлетворяющей им, была выбрана легковесная реляционная база данных SQLite; ее функционирование не требует отдельно работающего процесса, с которым взаимодействует программа. Электронные документы Царицынского городского магистрата конца XVIII в. хранятся в одном файле, а функционирование СУБД SQLite поддерживается самим приложением «Электронный магистрат».

Важным аспектом в процессе реализации приложения является организация модели структуры данных. Учитывая особенность рассматриваемой предметной области – разнородность атрибутивных характеристик документов, авторы программного продукта рассматривали в качестве перспективных следующие подходы к решению названной типовой задачи:

- осуществление проецирования вида документа на таблицу в базе данных (что использовано при разработке приложения «Электронный магистрат»). Для каждого вида документа при этом создается таблица, описывающая особенности группы, подгруппы, вида и разновидности документа. Преимуществом подобного подхода является простота и наглядность модели, интерфейса и логики приложения; в качестве недостатка можно назвать избыточность и повторяемость операций в процессе разработки и использования приложения;

- эмулярование связи «Ключ-Значение» в реляционной модели. Графическая нотация данного подхода представлена на рис. 1. (Document – таблица, в которой хранятся все регистрируемые документы ([...] — обозначает атрибуты, общие для всех видов); Attribute — справочник всех атрибутов; Document_Attribute — таблица, в которой хранятся значения для атрибута конкретного документа). Важным в рамках данного подхода является централизация всех данных системы в трех таблицах, возможность программного управления атрибутами документов; к основным недостаткам можно отнести сложность запросов на выборку и последующую их обработку, что, в конечном итоге, сказывается на производительности программы в целом и возникновении проблемы ее масштабирования; использование документно-ориентированной базы данных (Mon-goDB, Redis, MemcacheDB). Если предыдущий подход являлся лишь попыткой эмуляции решения «Ключ-Значение», то данная парадигма лежит в основе выше перечисленных систем и представляется наиболее подходящим вариантом. Однако ее использование существенно противоречит предъявляемым требованиям к приложению.

Интерфейсная часть приложения реализована с помощью языка высокого уровня программирования C#, инфраструктура которого обеспечивается средствами среды NET Framework 2.0. Поддержка среды осуществляется ОС Windows XP Service Pack 2 и выше.

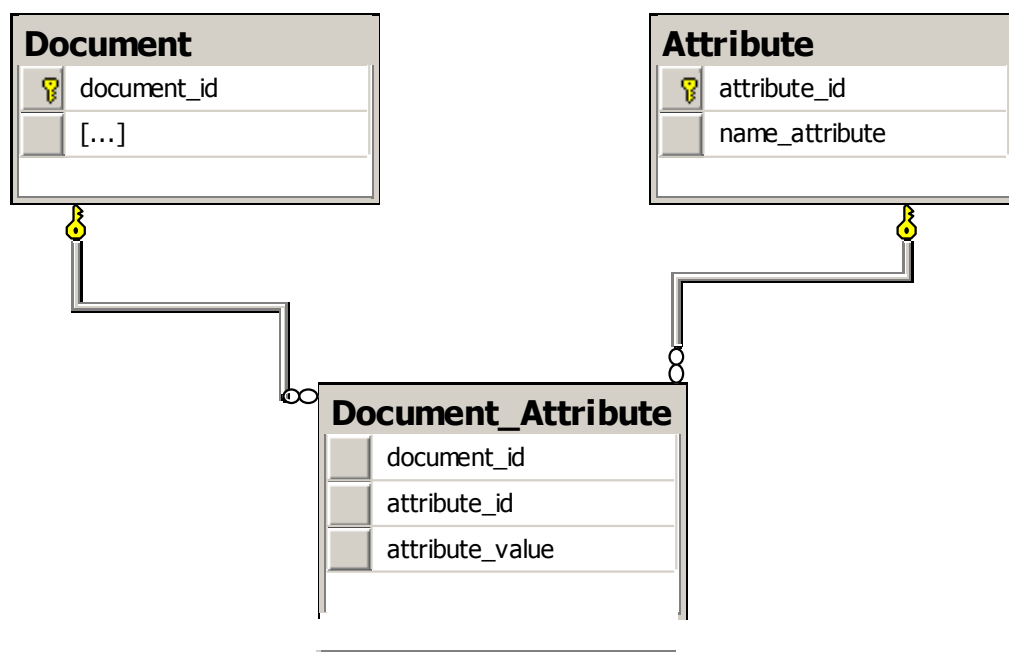


Рис. 1. Схема "Ключ-Значение" реляционной модели в нотации IDEF1x

Программный комплекс предназначен для учета, анализа структуры и функций документов, сохранившихся в составе архивного фонда Царицынского городского магистрата конца XVIII в.: книг и реестров, протокольной документации, различных видов отдельных документов (просительной, договорной, отчетной и др. документации).

«Электронный магистрат» позволяет просматривать внесенные документы, добавлять, редактировать и удалять их, осуществлять поиск на основе характеристик различных реквизитов (номера документа, даты поступления, даты решения по делу, наименования документа, подписи и т. д.), основных содержательных и структурно-композиционных параметров текста, целевого назначения бумаги.

Архивные тексты помещены в библиотеку документов. Первый уровень представлен группами «Книги и реестры», «Протокольная документация», «Раздельные документы». Второй и третий уровни включают работу с подгруппами, видами и разновидностями документов. Группа «Книги и реестры» содержит подгруппы «учетные книги» (книги прихода и расхода) «регистрационные документы» (реестры входящих указов и предложений, книги регистрации, книги записи), предназначенные для фиксации информации о поступающей и исходящей из магистрата корреспонденции, для регистрации договоров, прошений, вексельных обязательств жителей Царицынского округа. Группа «Протокольная документация» содержит подгруппы «информационно-справочные документы» (журналы) и «распорядительные документы (решения, которые созданы на основе журналов, содержащие постановление коллегиального органа — присутствия магистрата. Группа «Раздельные документы» состоит из подгрупп «распорядительные документы» (указы, приказы, наставления, предложения), «отчетные» (рапорты), «договорные» (верющие письма, контракты, купчие), «информационно-справочные» (выписки из дел, копии с журналов, сообщения, доношения, объявления), «просительные» (челобитные, прошения, жалобы), «акты принятия на себя каких-либо обязательств» (подписки, клятвенные обещания, расписки, одобрения), «записи показаний судебного и несудебного характера» (сказки, объяснения, допросы), «удостоверительные документы» (квитанции, расписки, паспорта), «учетные» (регистры, описи, ведомости).

Характеристика видов документов дается по атрибутам — структурно-содержательным признакам, составляющим формуляр и композиционную структуру текста деловой бумаги. С помощью функций «Добавить», «Редактировать», «Удалить», «Обновить» и «Просмотр» в электронную базу данных могут быть введены данные о новом документе (формуляре, композиционных частях и т. д.), отредактированы или аннулированы сведения об уже имеющихся единицах описания. Выделены обязательные и факультативные (непоследовательностью реализуемые) атрибуты документа. Так, для реестра входящих указов и предложений обязательными реквизитами формуляра являются название документа, а также графы таблицы, в которой заключено деловое содержание: № п/п, число регистрации, названия регистрируемых бумаг; к факультативным будут относиться даты принятия резолюции, исполнения дела, различные служебные отметки и т. д.. Композицию данного документного вида образует перечень обязательных сведений о поступивших в Царицынский городской магистрат указах и предложениях (название документа, адресант, краткое содержание и т. д.). «Электронный магистрат» позволяет загрузить сканированную копию текста или его фрагмента.

Разработанный на основе анализа архивного материала программный комплекс позволяет описать внутреннюю форму различных видов документов конца XVIII в., сделать выводы о степени ее устойчивости в зависимости от реализации системных и функциональных свойств в каждом конкретном случае, показать особенности правил составления официальных письменных текстов, способов систематизации информации в региональном делопроизводстве, что создает основу для реконструкции архивного фонда в его целостности.

Источник

Тупикова Н.А., Преферансов С.А., Вуйлов Д.А., Цельник Г.Л. Программный комплекс по описанию архивного фонда Царицынского городского магистрата (конец XVIII в.) // Свидетельство о государственной регистрации программы для ЭВМ № 2011617898 / Зарегистрировано в Реестре программ для ЭВМ 07.10.2011 / Роспатент: М., 2011.

Список литературы

Зубов, 2004 — Зубов А.В., Зубов И.И. Информационные технологии в лингвистике: Учеб. пособие. М., 2004.

Кибрик, 2001 — Кибрик, А.Е. Очерки по общим и прикладным вопросам языкознания. М., 2001.

Кузнецов — Кузнецов, С.Д. Проектирование и разработка корпоративных информационных систем. [Электронный ресурс] http://citforum.ru/cfin/prcorpsys/infosistpr_03.shtml.

Петров, 2002 — Петров В.Н. Информационные системы: Учебник. СПб., 2002.