

Набор и описание рукописей в «Технологии смешанного набора»

А. В. Коваленин, Е. Ю. Нечипоренко
«Фонд знаменных песнопений», Новосибирск, Россия

The principles of an electronic representation of manuscripts are discussed that can be suitable both for researchers' studies and for the maintenance of practical use of their texts by different groups of practitioners. The technique proposed for texts and descriptions of manuscripts helps combining different scripts and permits diversity of text markup interpretation. The approaches concerned are demonstrated with an example of the electronic corpus of ancient Russian singing manuscripts.

I. Постановка задачи

Певческие рукописи среди прочих древнерусских рукописей выделяются сегодня тем, что вновь находят непосредственного читателя, то есть не исследователя старины, а певчего. Устойчивость как литургического, так и интонационно-мелодического текста церковных песнопений такова, что стихира XV века сегодня может быть непосредственно исполнена в богослужении Русской Православной Церкви, где возрождается исконная певческая традиция. Это определило нашу практическую задачу — публикацию *текстов песнопений*, для чего был создан проект «Фонд знаменных песнопений» (далее — ЗФ)¹.

Такая формулировка задачи уже содержит в себе известную трудность. Общепринято теоретическое положение о том, что надо различать *текст* (то, что записано) и его *запись* (то, как записано). Строго говоря, публикуется не сам по себе текст, а какая-то его запись, из которой текст усматривается читателем. Практическая направленность нашего проекта сделала для нас важность этого положения очевидной. С одной (объективной) стороны, спо-

¹ «Знаменный фонд» [Электронный ресурс]. — Режим доступа: <http://znamen.ru>, свободный. — Загл. с экрана.

*Современные информационные технологии и письменное наследие:
от древних рукописей к электронным текстам*

соб записи одного и того же текста (графическая и орфографическая традиция) от рукописи к рукописи сильно отличается. С другой (субъективной) стороны, разные группы практиков жестко придерживаются своих традиций, в частности и потому, что непривычная запись затрудняет понимание текста и, тем самым, его практическое использование. Поэтому практическая публикация текста не может быть механическим воспроизведением знаков рукописи, а при выборе способа записи всегда ориентируется на конкретную рукописную или издательскую традицию.

Выход из этого затруднения возможен при электронном *хранении текста*, если под этим понимать хранение такой информации о нем, по которой можно воспроизвести его запись в разных традициях, а не только запись оригинала.

Конкретные вопросы, которые при этом возникают, обычно являются проявлением одной общей *проблемы восстановления текста по его записи* и относятся уже не к техническим, а к предметным (историческим, филологическим, литургическим) вопросам. Про эти вопросы мало сказать, что они не являются прерогативой наборщика корпуса: они не всегда разрешимы на основе материала набираемого песнопения, и даже материала целой рукописи, а порой так и остаются проблемными. Поэтому форма хранения текста должна быть рабочим инструментом исследователя, позволяя хранить и детали оригинальной записи, и необходимые орфографические варианты, и даже рабочие замечания исследователя.

Для постановки задачи немаловажной явилась и форма организации работы в ЗФ. Изначально предполагалось, что песнопения (а теперь и описания рукописей) набираются и присылаются в Фонд энтузиастами из разных мест. Для этого формат набора должен быть простым, а необходимый инструмент общедоступным.

II. «Технология смешанного набора»

«Технология смешанного набора» — это дисциплина работы с набором разных по назначению фрагментов. Певческие рукописи содержат в себе связанные между собой фрагменты разных письменностей — церковнославянского письма, древнерусской певческой нотации (крюков), нотолинейной записи мелодии. Их описания также содержат различные фрагменты: ссылки на песнопения и внетекстовые элементы, элементы различных разметок

рукописи, текстовые фрагменты, сделанные писцом, владельцем рукописи, редактором электронного корпуса.

Технология сводится к нескольким элементарным принципам.

1. Информация (песнопение или описание рукописи) хранится в простом «текстовом файле» (ASCII8), то есть набирается только терминальными символами и редактируется простейшими текстовыми редакторами.

2. *Представление знаков рукописи* не связано ни с какой товой кодировкой. Текстовый фрагмент любой письменности записывается в разработанном для нее терминальном представлении. Цель такого представления — сочетание точности сохранения деталей и относительной наглядности записи. Оно является и гибкой платформой для компьютерной обработки текста.

Для церковнославянского (древнерусского) письма за основу был взято принятое Сообществом «Славянская типографика»² в стандарте НІР представление знаков, которое мы дополнили для удобства работы с рукописями. Для крюковой и нотолинейной нотации мы разработали свои терминальные представления. Эти три представления уже вошли в практику как рабочие нотации для хранения текстов и их обсуждения в электронной переписке.

3. *Разметка*. Файл рассматривается как цепочка фрагментов. Фрагмент каждого рода предваряется специальным ключевым словом — меткой рода фрагмента. В файле не содержится никакой другой информации об интерпретации фрагментов.

Между фрагментами нет иерархических отношений. Каждой задаче (аспекту исследования) соответствует свое представление о структуре рукописи, которое можно не фиксировать в рабочем файле, а считать свойством конкретной интерпретации. В рабочем файле сосуществует множество независимых разметок.

4. *Интерпретация меток* фрагментов задается в отдельном («стилевом») файле, своем для каждой практической задачи. Она следует двум принципам.

4.1. Фрагмент, метке которого не предписана никакая интерпретация, *уничтожается*. Тем самым каждая задача (группа исследователей) имеет дело с одними фрагментами и игнорирует

² Сообщество «Славянская типографика» [Электронный ресурс]. — Режим доступа: <http://cslav.ru>, свободный. — Загл. с экрана.

другие. Стилиевой файл, таким образом, становится носителем взгляда на текст.

4.2. Интерпретация метки — это ее замена на последовательность, в которую может включаться цепочка имен содержательных преобразований фрагмента. В частности, таким преобразованием является перевод терминального представления в кодировку требуемого шрифта. Но большинство преобразований — такие, как смена регистра, раскрытие титл, снятие ударений, выбор орфографического варианта, — работает над терминальным представлением письменности. Чтобы такие преобразователи могли находиться в ведении специалистов-непрограммистов, создана возможность их реализации не только на алгоритмическом языке, но и в виде списков производимых специальным образом текстовых замен. Специалисты могут сами переопределять или дополнять системные преобразователи, создавая, например, собственные таблицы замен для чисто лингвистических алгоритмов, таких, как выделение приставок или простановка ударений. На практике большинство пользователей использует в работе готовые преобразователи и предписанные техническим редактором корпуса метки.

III. *Запись текстового фрагмента*

Запись текстового фрагмента содержит не только терминальное представление знаков рукописи, но и вспомогательные обозначения, необходимые для получения других вариантов записи, в том числе обязательно орфографически-нормативного варианта. Работающие с этими фрагментами информационные системы в зависимости от выбора пользователя указывают для их интерпретации различные цепочки преобразований. Цепочка преобразований, необходимая для получения записи оригинала, является атрибутом описания рукописи. Такой подход позволяет избежать повторения указаний на типичные особенности рукописи, которое всегда снижает наглядность рабочей записи. (Например, такой особенностью может быть отсутствие ударений или сплошное употребление «и» вместо «і».) Это открывает путь к формальному описанию орфографических традиций, для изучения которых по мере наполнения корпуса накапливается эмпирический материал.

IV. *Практические результаты*

В ЗФ сейчас доступно более 800 песнопений в формате Технологии смешанного набора, для которых при помощи специального

инструмента пользователи могут создавать наглядную форму (HTML), по своему усмотрению назначая такие параметры распечатки, как наличие того или иного слоя нотации (крюки, круглые или квадратные ноты), сокращенная и/или развернутая форма записи мелизматических фрагментов, наличие ударений, способ изображения междуслогового пространства в тексте и раздельно-речия, выбор оригинальной (как в рукописи) или нормативной орфографии. Рабочий инструмент ЗФ пока реализован на базе DOS-инструментов.

Перевод ЗФ на современную основу (на PHP) начат с описания рукописей и создания инфраструктуры работы с ними, с первоочередными функциями которой можно познакомиться на примере информационной системы «Певческие рукописи Красноярска»³. Система, по-разному интерпретируя файл описания рукописи, представляет пользователю разные ее стороны. В одном случае ссылки на песнопения перемежаются имеющимися в рукописи надписями, в другом — автоматически вставляемой рубрикой. Каждое из этих представлений допускает по усмотрению пользователя варьирование: текст показывается либо церковнославянским шрифтом (по выбору из шрифтов в двух разных кодировках), либо в рабочем представлении, либо в упрощенной (современной) письменности; в нормативной или оригинальной орфографии; с разной степенью детализации сведений о песнопениях. Над текстами описаний реализован поиск, позволяющий пользователю по упрощенному образцу находить в рукописях его вхождения во всех возможных орфографических вариантах.

Автоматическая вставка рубрикации опирается на специально разработанную систему шифров, отражающих систематическую классификацию богослужебных песнопений и указываемых во фрагментах-ссылках на песнопения. Наличие этих шифров позволяет организовывать запросы к описаниям из внешних программ.

Дальнейшее развитие ЗФ предполагает создание сводных оглавлений, порождение сводных текстов и перенос опробованных приемов (вариантность, преобразования, поиск) на работу с нотной и крюковой строками песнопений.

³ ИС «Певческие рукописи Красноярска» [Электронный ресурс]. — Режим доступа: <http://tuk.kraslib.ru>, свободный. — Загл. с экрана.

Подходы к построению древнерусского тезауруса (на материалах Словаря русского языка XI–XVII вв.)

О. В. Колобкова

Московский государственный технический университет имени
Н. Э. Баумана, Россия

The solutions of elaboration of thesauruses and linguistic knowledge bases are illustrated in the article. The texts of the Russian Language dictionary of XI–XVII centuries were used in the work as a source of terminological information. The system requirements and the methodological goals of investigation were also formulated.

Любая лексикографическая работа сопряжена с привлечением огромных ресурсов времени и труда лексикографов. Современные вычислительные средства дают возможность автоматизировать лексикографическую работу практически на всех этапах — от выбора цитат до редактирования словаря и его издания. В этой связи основным направлением предстоящего исследования будет поиск и создание инструментов, обеспечивающих высокую эффективность получения данных об исторической лексике древнерусского языка. Работа будет вестись на базе материалов Словаря русского языка XI–XVII вв. (далее — СЛРЯ).

СЛРЯ по жанру исторический, он основан на выборке лексики из памятников письменности древне- и старорусского периода. В список источников Словаря входят также «Словарь церковнославянского языка» А. Х. Востокова (т. I–II. СПб., 1858–1861 гг.) и «Материалы для Словаря древнерусского языка» И. И. Срезневского (т. I–III. М., 1958 г.; переиздание 1989 г.). Круг памятников ограничен хронологическими рамками, а для лексики XV–XVII вв. — принадлежностью русскому языку. За двадцать лет издано 27 выпусков, охватывающих отрезок от А до С. За это время сложились определенные традиции, приемы в подаче лексического материала. Сейчас этот словарь представляется наиболее пол-