Метаинформация в коллекции М. В. Ломоносова на портале «Манускрипт: Славянское письменное наследие»

Р. А. Аникина<sup>1</sup>, В. А. Баранов<sup>1</sup>, Т. В. Кокорина<sup>1</sup>, С. В. Ощепков<sup>2</sup>, А. А. Соколова<sup>2</sup>

<sup>1</sup>Ижевский государственный университет, <sup>2</sup>Удмуртский государственный университет, Ижевск, Россия

Работы коллектива проекта «Манускрипт» (http://manuscripts.ru/) по созданию полнотекстовой базы данных не ограничиваются подготовкой электронных коллекций рукописей древнейшего и средневекового периода славянской письменности. В настоящее время на основе разработанной информационно-аналитической системы «Манускрипт» и ее модулей, обеспечивающих различные этапы подготовки и визуализации данных, создается коллекция произведений М. В. Ломоносова. Основой для коллекции является Полное собрание сочинений в 10 томах, изданное в 50-х гг. Академией наук СССР<sup>1</sup>.

Несмотря на автоматизацию некоторых этапов подготовки коллекции – сканирование печатного текста, его распознавание, загрузка в базу данных, основная часть работы – сверка с печатным изданием, исправление ошибок автоматического ввода, тегирование, обеспечивающее, например, сохранение информации о таблицах и о графических компонентах печатного издания, – остается чрезвычайно трудоемким и долговременным процессом.

Не менее сложным этапом подготовки коллекции является ввод метаинформации о произведениях. Метаинформация в настоящее время используется при поиске и отборе текстов коллекции для последующей подготовки справочных материалов – выборок, конкордансов, сравнительных указателей.

В связи с ориентацией системы в первую очередь на лингвистический анализ поля предназначены в основном для характеристики произведений, но не печатного издания.

Две задачи были решены на этапе ввода метаданных: (1) определение перечня характеристик и (2) установление стандарта текстовых полей – состава информации поля, порядка ее следования, приемов сокращения и некоторые другие.

Особенностью формата значений метахарактеристик является, в частности, ориентация на поисковые функции и на их реализацию в web-модуле коллекции, поэтому в настоящее время (1) значения, для которых существуют общепринятые сокращения, даются в поле свойства полностью, (2) в одном поле могут быть значения различных зон библиографического описания, обычно помещаемые в базах данных в различных полях. Это позволяет обеспечить поиск как по полному наименованию значения, так и по его части, а также не дробить поля.

<sup>&</sup>lt;sup>1</sup> Ломоносов М. В. Полное собрание сочинений : в 10 тт. / М. В. Ломоносов. М. ; Л. : Изд-во АН СССР, 1950–1959; Т. 11. Л. : Наука, 1984.

Перечислим поля, используемые в первой версии коллекции произведений М. В. Ломоносова, укажем их форматы и приведем примеры значений (URL: http://manuscripts.ru/mns/portal.main?p1=31)<sup>2</sup>:

## Метаданные печатного издания:

- Автор: [Фамилия] <sup>3</sup> [Имя] [Отчество] *Ломоносов Михаил Васильевич.*
- Заглавие печатного издания: [Автор И. О.] Название издания [. Т., вып.] Ломоносов М. В. Полное собрание сочинений. Т. 10.
- Продолжение заглавия: [Продолжение заглавия] *Служебные документы и письма, 1734-1765 гг.*
- Место издания: Наименование населенного пункта1 [ Наименование населенного пункта2] – Москва ; Ленинград.
- Издательство: [Название издательства] Издательство Академии Наук СССР.
- Год издания: гггг[–гггг] 1957.
- Условное название: Сокращенное название коллекции.
  Сокращенное название издания ЛМН. ПСС, 10.

## Метаданные произведения:

Название произведения: «Начальная синтагма...» [Авторское / научное заглавие] – «Бугристы берега, благоприятны влаги...» [О сомнительном произношении буквы Г в российском языке].

## Примечания:

- 1. Идентификация произведения осуществляется по его первой синтагме; заглавие помещается после него в квадратных скобках.
- 2. Авторское заглавие в случае его большой длины может быть сокращено; пропуск отмечается многоточием.
- 3. При отсутствии авторского или научного названия / заглавия для идентификации произведения используется только его начальная синтагма: «Начальная синтагма...» «Богиня, Дщерь Божеств, науки основавших...».
- 4. В случае недостаточной уникальности первой синтагмы (например, формульность начала текста) для идентификации произведения используются последующие фразы, при этом уникальные слова пропущенных формул заключаются в квадратные скобки: [уникальные словоформы/словосочетания формулы] «Следующая синтагма...» [...рейхсграф... Роман Ларионович!] «Не имея довольного случая изъяснить перед вашим сиятельством...».
- Автор: [Фамилия] [И.] [O.] *Ломоносов М. В.*
- Порядковый номер: [порядковый номер в издании] *17*.
- Жанр: [название жанра] поэзия.

<sup>2</sup> Приводится название свойства описываемого объекта, через двоеточие – формат значения, через тире курсивом – пример значения.

<sup>3</sup> В квадратные скобки заключены маски значений, которые могут не присваиваться объекту.

- Место создания: [Наименование населенного пункта] Санкт-Петербура.
- Время создания: [дд].[мм].гггг [– [дд].[мм].гггг] 19.11.1747.
- Переводность: да / нет нет.
- Аутентичность: оригинал / копия оригинал.
- Язык: [язык] русский.
- Условное название: Сокращенное название коллекции.
  Сокращенное название издания, номер произведения в томе ЛМН. ПСС, 10, 13.
- Тема произведения: [тема произведения] благодарности.
- Субъекты произведения: [Имя] Тредиаковский В. К.
- Номер начальной страницы: [арабские цифры] 344 и некоторые другие (интерфейс для ввода и редактирования данных показан на рис. 1).

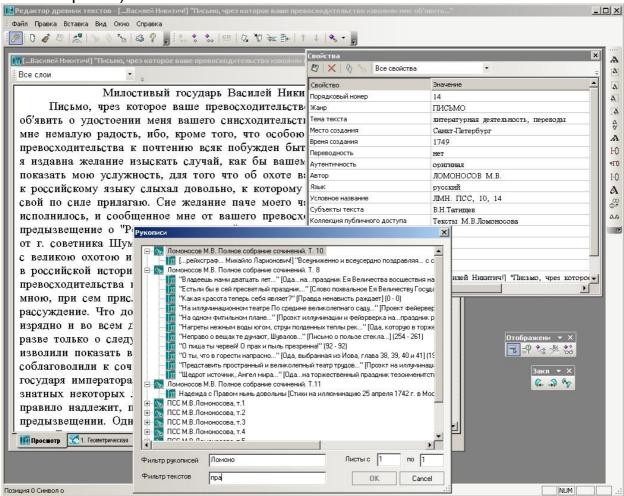


Рис. 1. Отбор произведений для ввода метаинформации и редактирование данных в редакторе OldEd

Для хранения и ввода характеристик предусмотрено использование справочников, в частности справочников *Люди* и *События*.

Поиск и отбор текстов на основе метаданных осуществляется с помощью специализированного web-модуля, имеющего два режима работы — простой (URL: http://manuscripts.ru/mns/srch.simple?p\_ed\_id=50584966) и расширенный (URL: http://manuscripts.ru/mns/srch.complex?p\_lang=RU&p\_ed\_id=50584966).

Первый на основе введенной текстовой маски искомого значения обеспечивает поиск томов и текстов, вывод перечня единиц, обладающих этим значением и показ свойств, которым принадлежат эти значения

(рис. 2). Поиск по мета описаниям - Mozilla Firefo: \_ | | ×  $\underline{\Phi}$ айл Правка <u>В</u>ид <u>Ж</u>урнал <u>З</u>акладки <u>И</u>нструменты <u>С</u>правка 🖚 🕶 🔻 🕝 🚳 👔 👪 http://manuscripts.ru/mns/srch.search ▼ ▶ G Google 👪 Поиск по мета описаниям Расширенный поиск Мои выборки RU EN 3anpoc [?]: 174% искать □ Искать в найденном □ Ломоносов М.В. Полное собрание сочинений. Т. 10 [...Василей Кирилович!] "Вы изволили мне об'явить словесно..." Скрыть свойства <u>Время создания</u>: 1748 [...Василей Никитич!] "Письмо, чрез которое ваше превосходительство изволили мне об'явить..." Скрыть свойства Время создания: 1749 [...Григорей Николаевич!] "Хотя должность моя и требует, чтобы по присланному..." [...Иван Антонович!] "Присланные от вашего превосходительства полфунта..." Показать свойства [...граф... Кирил Григорьевич!] "Ведая, что ваше сиятельство имеете к наукам любовь и ревность..." Показать свойства [Высокоблагородный г. советник...] "В присланном от вашего высокоблагородия ко мне..." □ Ломоносов М.В. Полное собрание сочинений, Т. 8 "Ах, жалко! Без тебя лилеи почернели... Скрыть свойства Время создания: 1747 "Ах, когдаб погрязнул в море чужелобник..." Скрыть свойства Время создания: 1747 "Без всякой мы вины любовь зовем слепою..." "Благословен Господь мой Бог..." [Преложение псалма 143] "Богиня красотой, породой Ты богиня..." [Надпись на иллуминацию...в день тезоименитства Ея Величества 1748 года...] Показать свойства "В сей день, блаженная Россия, любезна небесам страна..." [Ода на день рождения Ея Величества Государыни Императрицы...] "Взирает верность на Тебя во храме полнаго блаженства...

Рис. 2. Отбор произведений в режиме простого поиска на основе маски метаданных

Второй дает возможность сформировать запрос на основе нескольких параметров: выбора единицы (печатного издания / произведения / фрагмента), ее свойств и значений свойств. Динамичное подгружение значений по мере ввода маски позволяет просмотреть, выбрать и указать значение, необходимое для выборки.

При построении запроса в модуле расширенного поиска могут быть использованы значения нескольких параметров одной единицы или несколько параметров различных единиц. При подготовке такого запроса используются логические  $\mathcal{U}$  или  $\mathcal{U}\mathcal{N}\mathcal{U}$ , а также логическое  $\mathcal{H}\mathcal{E}$ , исключающее из результата поиска единицы с указанным значением (рис. 3).

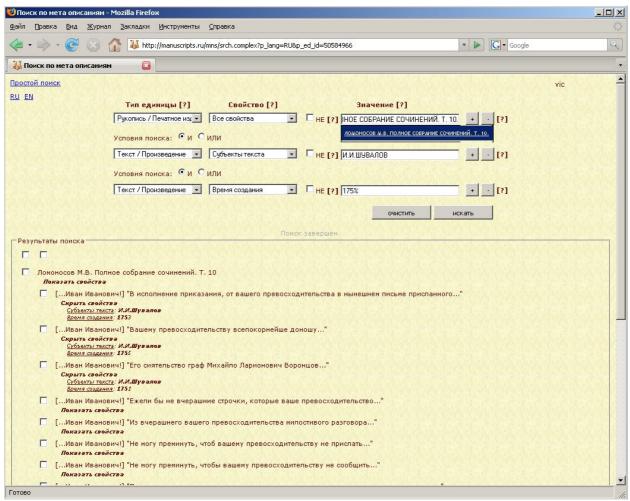


Рис. 3. Отбор произведений в режиме расширенного поиска на основе нескольких параметров

В целом хранимая в базе данных «Манускрипт» метаинформация о произведениях М. В. Ломоносова обеспечивает гибкий отбор текстов для последующей подготовки справочных материалов в лингвистических и лингвотекстологических исследованиях.

## Благодарности

Работа выполняется в рамках проекта, поддержанного Российским гуманитарным научным фондом (РГНФ), проект № 07-04-12147в.

Meta-information in the Lomonosov collection in the Manuscript portal Regina A. Anikina<sup>1</sup>, Victor A. Baranov<sup>1</sup>, Tatyana V. Kokorina<sup>1</sup>, Sergey V. Oshchepkov<sup>2</sup>, Anastasiya A. Sokolova<sup>2</sup>

<sup>1</sup>Izhevsk State Technical University, <sup>2</sup>Udmurtia State University Izhevsk, Russia This paper discusses the experience of preparing an electronic fulltext collection of the works of Lomonosov. Attention is paid to questions of metadata creation, the choice of fields, and their formats.