

МИНИСТЕРСТВО НА ОБРАЗОВАНИЕТО И НАУКАТА НА РЕПУБЛИКА БЪЛГАРИЯ
КИРИЛО-МЕТОДИЕВСКИ НАУЧЕН ЦЕНТЪР ПРИ БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. М. Т. КАЛАШНИКОВА
НАУЧНОЕ СООБЩЕСТВО “ПИСЬМЕННОЕ НАСЛЕДИЕ”
DIGITAL MEDIEVALIST SCHOLARLY COMMUNITY
ФОНДАЦИЯ „УСТОЙЧИВО РАЗВИТИЕ НА БЪЛГАРИЯ“

**Писменото наследство
и информационните технологии**

El’Manuscript–2014

Материали от V международна научна конференция
Варна, 15–20 септември 2014 г.

София · Ижевск
2014

Сборникът е издаден с финансовата подкрепа на Министерството на образованието и науката на Република България по процедура за подкрепа на международни научни форуми.

Отговорни редактори: проф. д-р В. А. Баранов
 доц. д-р В. Желязкова
 д-р А. М. Лаврентъев

Редактори: Нели Ганчева, Веселка Желязкова (български текст)
 О. В. Зуга, В. А. Баранов (руски текст)
 Кевин Хокинс (Kevin Hawkins) (английски текст)

Писменото наследство и информационните технологии [Текст] : материали от V международна науч. конф. (Варна, 15–20 септември 2014 г.) / отг. ред. В. А. Баранов, В. Желязкова, А. М. Лаврентъев. — София ; Ижевск, 2014. — 448 с.

Сборникът съдържа материали от конференция, посветена на разработването и създаването на съвременни средства за съхраняване, описване, обработка, анализ и публикуване на ръкописни и старопечатни книжовни паметници и исторически извори, а също и на въпросите за подготвянето на електронни ресурси в областта на хуманитаристиката и тяхното използване в научните изследвания и преподаването.

© Кирило-Методиевски научен център — БАН, 2014
© Ижевский государственный технический университет
им. М. Т. Калашникова, 2014
© Авторски колектив, 2014
© Лилия Тошкова — графичен дизайн на корицата, 2014

ISBN 978–954–9787–25–2

Научные и учебные проекты на базе исследований Словаря Академии Российской 1789–1794 гг.

А. Ю. Филиппович

Словарь Академии Российской, автоматизация исторических исследований, распознавание, факсимильные шрифты, учебные проекты, информационная технология переиздания

Scientific and Educational Projects based on Research on the Russian Academy Dictionary of 1789–1794

Anna Philippovich

The report considers research based on the reissue of the Russian Academy Dictionary of 1789–1794 (RAD) and two grand-funded follow-on projects the creation of an integrated toolkit for the automation of research on the RAD, and the development of vocabulary components of an integrated information technology for the reissue of 18th– and early 19th-century printed sources. Their main objectives and results are described. Examples of educational projects in font creation, frequency studies of historical texts and OCR technology are provided.

Проект переиздания

Отправной точкой исследований, описанных в данном докладе, стал проект переиздания *Словаря Академии Российской 1789–1794 гг.* [САР], который является уникальным памятником русской исторической лексикографии.

Проект переиздания *Словаря Академии Российской 1789–1794 гг.* — это печатное “издание факсимильного типа с исправлениями и дополнениями” [Филиппович Ю. 2001: 8], с максимально возможной полиграфической точностью воспроизводящее подлинник. При этом была выбрана наборная форма переиздания, где “традиционные этапы набора и корректуры текста дополняются разработкой факсимильной шрифтовой гарнитуры, создаваемой на основе старинной оригинальной” [Филиппович А. 2008: 68]. Наборная форма книги позволила сформировать основу для создания словарной базы данных и электронного издания. Таким образом в результате реализации этого проекта были начаты исследования в различных научных областях, в том числе филологии, истории, лексикографии, технологии печати, шрифтографии, в области информационных технологий и электронного издательства. Уникальная технология переиздания САР объединила ученых и специалистов из ведущих научно-исследовательских институтов и университетов в том числе: МГТУ им. Н. Э. Баумана, МГУП имени Ивана Фёдорова, Института Языкознания РАН, Института Русского языка РАН

МГИ им. Е. Р. Дашковой. В данном докладе мы рассмотрим проекты в области информационных технологий.

Научно-исследовательские проекты

В продолжении начатых исследований при поддержке РГНФ в 2006–2008 годах был реализован проект “Интегрированная инструментальная информационно-программная среда для автоматизации исследований Словаря Академии Российской 1789–1794 гг.” (№ 06–04–12412в). Созданная среда включала следующие компоненты: лингвистическую базу данных; программный комплекс автоматизированного ввода текста словарных статей в базу данных; программный комплекс создания электронных словарей (прямых и обратных, частотных, заголовочных слов, словоформ эксцерпций и дефиниций), словоуказателей (общих, предметных, именных), словарных подмножеств (языковых и тематических), поисковых тезаурусов; двухуровневую (запросную и гипертекстовую) информационно-поисковую систему; биобиблиографическую информационную систему создателей Словаря; гиперграфическую систему факсимильных копий страниц оригинального издания Словаря. Подробнее смотри [Альманах].

Далее последовал проект “Разработка словарных компонентов интегрированной информационной технологии переиздания печатных источников XVIII – нач. XIX вв.” (МК-3732.2010.9).

Представляемая технология основана на выделении группы исторических источников с определенными свойствами (в числе которых способ печати, используемые средства оформления, шрифтовые гарнитуры и т. п.) и обработке их системами распознавания и индексирования текстов. Основная идея проекта: необходимо сформировать в электронной форме лексическое ядро языка коллекции документов.

Материалами для исследований выступали печатные источники XVIII – нач. XIX вв. В качестве основы ядра использовался САР.

Основные задачи проекта:

1. Анализ печатных источников XVIII – нач. XIX вв. и выявление их палеографических и лексических характеристик.
2. Исследование современных технологий оцифровки книг, методик ввода и обработки текстов и изображений, OCR-систем, ИПС и систем автоматического индексирования документов.
3. Создание лексического ядра коллекции документов XVIII – нач. XIX вв. на основе БД САР.
4. Разработка информационной технологии переиздания источников XVIII – нач. XIX вв. и исследование ее эффективности.
5. Подготовка учебно-научных материалов для исследования эффективности представленной технологии и подготовки ее отдельных компонентов.

Результаты проекта представлены в работах [Филиппович, Черкасова 2009; Филиппович А. 2010а; Филиппович А. 2010б; Филиппович 2012].

Учебные проекты

Тесная интеграция учебного процесса и научно-исследовательской работы в высшей школе создает предпосылки для формирования учебных проектов на базе научных исследований. Необходимым требованием, сформированным в современных учебных стандартах и программах обучения для специалистов старших курсов, магистров и аспирантов является решение научно-исследовательских задач. Такие задачи были решены в различных учебных проектах на базе исследований САР.

В рамках созданной технологии переиздания САР введено понятие “факсимильный шрифт” и разработана методика его создания [Филиппович 2013]. На базе этих разработок и исследований были выполнены ряд учебных проектов в области шрифтографики. Были разработаны практические задания, лабораторные работы и задания для самостоятельной работы по созданию элементов факсимильных шрифтов и орнаментов в программах-конструкторах шрифтов. Данные задания были ориентированы на студентов технических специальностей и проводились в рамках дисциплины “Лингвистическое обеспечение АСОИУ”, читаемой на 5-ом курсе в МГТУ им. Н. Э. Баумана.

В других учебных проектах необходимо было проанализировать факторы, влияющие на эффективность распознавания исторических текстов с помощью современных OCR-систем на основе количественных исследований. Среди них домашнее задание “Исследование системы оптического распознавания текстов” и курсовая работа на тему “Информационная технология исследования языка печатных источников XVIII – нач. XIX вв.”.

Ряд тем были сформированы в рамках индивидуальных заданий и квалификационных работ, в результате были созданы различные индексы на основе САР, словоуказатели, компоненты базы данных, различные проекции и т. д.

Таким образом, проект переиздания Словаря Академии Российской 1789–1794 гг. и начатые на его основе исследования коснулись более тысячи студентов, что позволило сформировать специалистов, способных работать в смежных научных областях, интегрировать гуманитарную и техническую сферу.

Литература

- Альманах — *Альманах исследований Словаря Академии Российской* [Электронный ресурс]. Режим доступа: <http://it-claim.ru/Projects/ESAR/Almanah.htm>, свободный.
- САР — *Словарь Академии Российской 1789–1794*. Т. 1–6. М.: МГИ им. Е. Р. Дашковой, 2001–2005.
- Филиппович А. 2008 — *Филиппович А. Ю.* Словарь Академии Российской (1789–1794): информационная технология переиздания. Вступительная статья М. И. Чернышевой. М.: МГУП, 2008.
- Филиппович А., Черкасова 2009 — *Филиппович А. Ю., Черкасова Г. А.* Книжное и электронное переиздание “Словаря Академии Российской” 1789–1794 гг. // Наука

- о книге: Традиции и инновации: К 50-летию сборника “Книга. Исследования и материалы”. М.: Наука, 2009. С. 361.
- Филиппович А. 2010а — *Филиппович А. Ю.* Разработка словарных компонентов интегрированной информационной технологии переиздания печатных источников XVIII – начала XIX века // Печатные средства информации в современном обществе (к 80-летию МГУП): Материалы науч. межвуз. конф. Сб. тезисов докладов. М.: 2010. С. 82–85
- Филиппович А. 2010б — *Филиппович А. Ю.* Информационная технология формирования лексического ядра языка печатных источников XVIII – нач. XIX вв. на основе Словаря Академии Российской 1789–1794 гг. // Горизонты прикладной лингвистики и лингвистических технологий [Электронное издание: компакт-диск]: Доклады междунар. науч. конф. Украина, Крым, Партенит. Симферополь, 2010.
- Филиппович А. 2012 — *Филиппович А. Ю.* Частотные исследования текстов источников гражданской печати XVIII века // Информационные технологии и письменное наследие: материалы IV междунар. науч. конф. Петрозаводск; Ижевск, 2012. С. 273–274.
- Филиппович А. 2013 — *Филиппович А. Ю.* Технология создания факсимильной шрифтовой гарнитуры для переиздания исторических источников // Проблемы полиграфии и издательского дела. 2013. № 4. С. 102–112.
- Филиппович Ю. 2001 — *Филиппович Ю. Н.* О переиздании Словаря Академии Российской 1789–1794 // Словарь Академии Российской 1789–1794. Т. 1. М.: МГИ им. Е. Р. Дашковой, 2001.