

Филиппович Анна Юрьевна

Интегрированная информационно-программная среда для автоматизации исследований Словаря Академии Российской 1789–1794 гг.*¹

На сегодняшний день широкое распространение среди электронных лексикографических ресурсов получили современные переводные, энциклопедические и тематические словари, в то время как большинство старинных источников является недоступным. Решением проблемы доступности – введение источников (книг) в научный оборот – является их копирование и переиздание. Использование компьютерной техники в полиграфии и распространение Интернет-технологий служит основой для создания электронных изданий и информационно-программных систем наряду с традиционными печатными.

Представленная работа посвящена созданию инструментальных средств для автоматизации исследований и анализа уникального памятника русской лексикографии Словаря Академии Российской 1789–1794 гг. (САР). Предпосылкой для данной разработки послужила подготовка печатного переиздания словаря, вышедшего в 2001–2006 гг. Интегрированная информационно-программная среда (ИПС) для автоматизации исследований САР предназначена для исследователей, изучающих языковую культуру XVIII века, филологов, лексикографов, историков, издателей словарей, преподавателей, аспирантов и студентов, работающих в этой области.

Интегрированная ИПС состоит из шести компонент: 1) лингвистическая база данных; 2) программный комплекс автоматизированного ввода текста словарных статей в базу данных; 3) программный комплекс создания электронных словников, словоуказателей, словарных подмножеств, поисковых тезаурусов; 4) двухуровневая информационно-поисковая система; 5) библиографическая информационная система «Создатели САР»; 6) гиперграфическая система факсимильной копии оригинального издания. В результате создается электронное издание САР, которое может быть доступно для широкого использования.

Представлена технология создания лингвистической базы данных САР. Основные ее проблемы связаны со сложной организацией словаря, его гнездовой структурой, ошибками и различиями в шрифтовой разметке, а также особенностями компьютерного представления исходных текстовых данных САР (формат, стилевая разметка, шрифты и т.д.). Рассмотрена технология обработки изображений страниц для формирования гиперграфической системы факсимильной копии оригинального издания, основные недостатки изображений и требования; представлен макет библиографической информационной системы «Создатели САР».

Разработка информационно-программной среды, печатного и электронного изданий САР, создание технологий и методов служит основой для переиздания других книг XVIII века и старинных печатных лексикографических изданий.

* Работа выполняется при поддержке гранта № 06-04-12412в Российского гуманитарного научного фонда и МГТУ им.Н.Э.Баумана Грант (2006–2008 гг.).