

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ФОРМИРОВАНИЯ ЛЕКСИЧЕСКОГО ЯДРА ЯЗЫКА ПЕЧАТНЫХ ИСТОЧНИКОВ XVIII – НАЧ. XIX ВВ. НА ОСНОВЕ СЛОВАРЯ АКАДЕМИИ РОССИЙСКОЙ 1789-1794 гг.*

Филиппович Анна Юрьевна

*Московский государственный технический университет им. Н.Э. Баумана,
кафедра «Системы обработки информации и управления» (ИУ5), Москва, РФ.
anna@philippovich.ru*

На протяжении всего существования цивилизации научное сообщество создает множество источников знаний – книг, и проблема сохранения и обеспечения доступа к ним исследователей научного и культурного наследия не теряет своей актуальности сегодня. Для решения этой проблемы сформулирована задача перевода источников и архивных документов в цифровую форму. Исторические документы переносятся на новые носители информации, создаются их факсимильные копии, которые сопровождаются библиографическими описаниями. Такие описания не раскрывают содержание документа, что ограничивает поисковые запросы исследователей. Т.о. актуальной задачей современности является решение проблемы поиска информации в созданных электронных хранилищах документов по их содержанию.

В докладе предлагается реализация этой задачи, основанная на формировании информационной технологии, суть которой в выделении группы 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. Основным компонентом таких систем является лингвистическая база данных, ядро которой – лексическая система языка рассматриваемого исторического периода. Т.о. первоначально необходимо сформировать в электронной форме лексическое ядро языка коллекции документов. Материалами для исследований выступают письменные источники XVIII – нач. XIX вв. В качестве основы ядра используется Словарь Академии Российской 1789-1794 гг., содержащий более 200000 лексических единиц.

На основе созданного лексического ядра и других словарных компонент могут быть разработаны информационно-поисковые системы, реализующие эффективный доступ к информации, хранящейся в источниках XVIII – нач. XIX вв. Это обеспечит информационным материалом и современным инструментарием ученых, занимающихся проблемами формирования и развития норм русского литературного языка и повысит эффективность проведения исследований в области исторической лексикологии и лексикографии.

* Работа выполняется при поддержке Гранта Президента РФ МК-3732.2010.9 (2010-2011 гг.).