

**План-проспект учебно-научного издания
«Методы оцифровки и распознавания
рукописных и первопечатных древнерусских источников»**

Аннотация

Настоящая книга посвящена решению проблемы сохранения культурно-исторического наследия. Она содержит научные и учебные материалы по современным технологиям переиздания и оцифровки рукописных и печатных древнерусских источников. К книге приводятся результаты теоретических и практических разработок в области распознавания старинных текстов на примере скорописи XVII в. и печатных источников XVIII – нач. XIX в.

Книга адресована ученым и специалистам, занимающимся распознаванием и оцифровкой древнерусских источников, студентам технических и гуманитарных специальностей. Она может быть использована в качестве учебного издания для подготовки бакалавров, обучающихся по направлению 230400 – «Информационные системы и технологии» (профили: «Информационные технологии в медиаиндустрии», «Информационные технологии в дизайне») и магистров обучающихся по программе «Компьютерная лингвистика и семиотика» направлений 230100 «Информатика и вычислительная техника» и 230400 – «Информационные системы и технологии».

Исследования и разработки выполняются при поддержке гранта Президента РФ № МК-3732.2010.9 в рамках конкурса молодых российских ученых-кандидатов наук 2010-2011 гг.

Краткое содержание книги

ВВЕДЕНИЕ

Раздел 1. ОСОБЕННОСТИ РУКОПИСНЫХ И ПЕЧАТНЫХ ДРЕВНЕРУССКИХ ИСТОЧНИКОВ

- 1.1. Краткая история славянского кирилловского письма
- 1.2. Палеография рукописных древнерусских источников
- 1.3. Особенности печатных древнерусских источников

Раздел 2. ТЕХНОЛОГИИ ОЦИФРОВКИ ИСТОЧНИКОВ

- 2.1. Факсимильные переиздания
- 2.2. Наборные переиздания

Раздел 3. РАСПОЗНАВАНИЕ РУКОПИСНЫХ ИСТОЧНИКОВ

- 3.1. Методы распознавания образов
- 3.2. Распознавание рукописного текста (на примере скорописи XVII в.)

Раздел 4. РАСПОЗНАВАНИЕ ПЕЧАТНЫХ ИСТОЧНИКОВ

- 4.1. Системы оптического распознавания
- 4.2. Распознавание источников XVIII – нач. XIX вв.

ЗАКЛЮЧЕНИЕ

ЛИТЕРАТУРА

УЧЕБНО-СПРАВОЧНЫЕ МАТЕРИАЛЫ

- Словарь терминов
- Список сокращений
- Библиотека
- Практические занятия
- Примеры курсовых работ и домашних заданий
- Вопросы для самопроверки и контроля знаний

ПРИЛОЖЕНИЕ (материалы исследований)

Объем и структура книги

Книга состоит из введения, четырех разделов, заключения, литературы, учебно-справочных материалов и приложения.

Объем книги составляет ок. 12 авт.л. (200-256 печатных страниц). Объем первой части – 20 страниц, остальных примерно одинаков – по 40 страниц. Вступительная и заключительная части, в которые входят введение, заключение, литература составляет 30 страниц, учебно-справочные материалы – 40-50 страниц, приложение – 30-40 страниц.

Описание содержания книги

Во *введении* представлены актуальность проблемы, рассматриваемой в книге, описание исследований, на которых базируются материалы книги, структура книги.

Первый раздел книги «Особенности рукописных и печатных древнерусских источников» посвящен вопросам истории славянской письменности, технологиям и характеристикам рукописных, первопечатных и далее старопечатных источников вплоть до XIX в.

Второй раздел «Технологии оцифровки источников» описывает особенности методов переиздания старинных печатных и рукописных книг. Рассматриваются две основные технологии: факсимильного и наборного переиздания. Описаны этапы обработки изображений, форматы данных, используемые для создания печатных и электронных переизданий, приведены примеры. Представлены методики создания факсимильных шрифтов, особенности набора и корректуры старинных текстов.

Третий раздел «Распознавание рукописных источников» включает рассмотрение теоретических основ распознавания образов и технологии распознавания скорописи XVII в. Рассматриваются типы задач и способы описания объектов распознавания, виды процедур предъявления обучающих выборок и правил классификации, формальные математические постановки задач, примеры систем их применение. Далее приводится описание технологии распознавание скорописных текстов, которое включает анализ методов распознавания, применительно к данной задаче. Представлена математическая постановка задач, описаны этапы проектирование системы распознавание и ее компоненты.

Четвертый раздел «Распознавание печатных источников» включает описание современных систем оптического распознавания печатного текста (OCR-систем), принципов и методов их построения. Приведены примеры современных OCR-систем. Далее рассмотрена технология распознавание источников XVIII – нач. XIX вв. с помощью OCR-системы: основные этапы, методики подготовки источников, анализ эффективности использования.

В *«Заключении»* приводятся основные выводы и рекомендации по проведенным исследованиям.

Раздел «Учебно-справочные материалы» содержит материалы для обучения студентов технических специальностей: описание практических занятий по теме распознавания текстов, примеры курсовых работ и домашних заданий, вопросы для самопроверки и контроля знаний. Библиотека содержит аннотированный тематический список литературы. Даны справочные материалы: словарь основных терминов, список сокращений.

Приложение содержит материалы исследований, лежащих в основе издания.