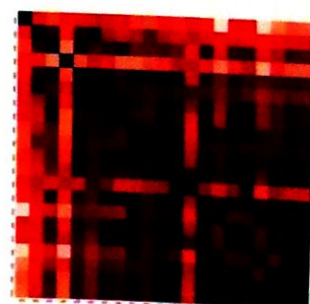
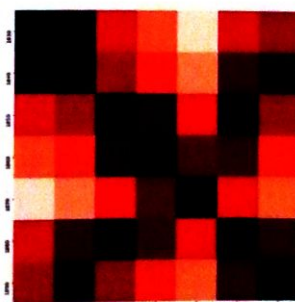
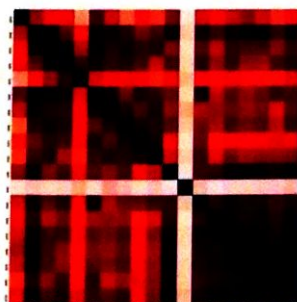
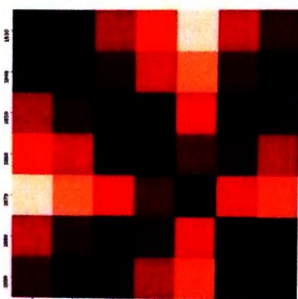
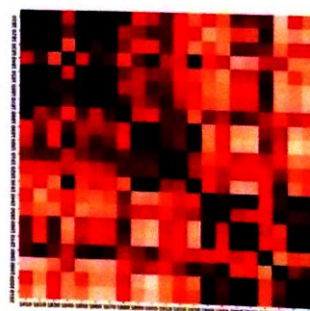
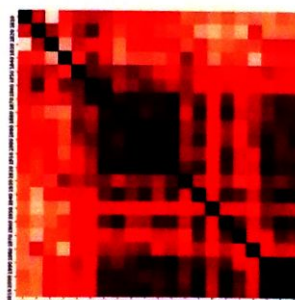
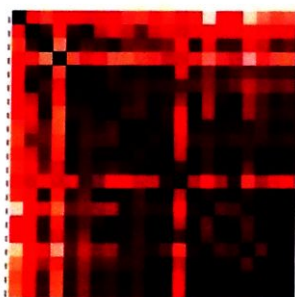
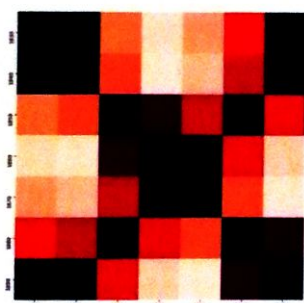


Министерство просвещения Российской Федерации
ФГБОУ ВО «Ярославский государственный педагогический
университет им. К. Д. Ушинского»

Бойчук Е.И., Лагутина Н.С., Лагутина К.В., Воронцова И.А.,
Шляхтина Е.В., Мишенькина Е.В., Беляева О.В.

АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ ТЕКСТОВЫХ ХАРАКТЕРИСТИК



УДК 004
ББК 32
А22

*Исследование выполнено при финансовой поддержке РФФИ
в рамках научного проекта №19-07-00243*

Рецензенты:

В.А. Соколов – доктор физико-математических наук, профессор, заведующий кафедрой теоретической информатики ФГБОУ ВО «Ярославский государственный университет им. П. Г. Демидова».

Т.Н. Хомутова – доктор филологических наук, профессор, заведующий кафедрой лингвистики и перевода института лингвистики и международных коммуникаций ФГАОУ ВО «Южноуральский государственный университет (НИУ)».

Бойчук Е.И., Лагутина Н.С., Лагутина К.В., Воронцова И.А., Шляхтина Е.В., Мишенькина Е.В., Беляева О.В.

А22 Автоматизированный анализ текстовых характеристик : учебное пособие / Е.И. Бойчук, Н.С. Лагутина, К.В. Лагутина, И.А. Воронцова, Е.В. Шляхтина, Е.В. Мишенькина, О.В. Беляева. – Ярославль, 2020. – 88 с.

ISBN 978-5-91730-998-9

Данное учебное пособие ориентировано на студентов филологических факультетов и факультетов информатики и вычислительной техники, увлекающихся автоматизированной обработкой текстов различных жанров. В пособии представлены некоторые формы работы с текстом с опорой на приложения и программы, облегчающие исследовательскую работу с материалом, его количественной и качественной обработкой. Представленные приложения и методики ориентированы на работу с русским языком, а также с различными иностранными языками, в частности английским, французским, испанским. Среди прочих представлено приложение, позволяющее проводить ритмический анализ текстов в четырех иностранных языках (ProseRhythmDetector). Данное приложение основано на понимании ритма как периодической повторяемости определенных языковых элементов в тексте.

Структура пособия включает пять глав. В первой раскрываются возможности компьютерной обработки текстов разных типов и жанров в родном и иностранном языках. Во второй главе представлен обзор различных корпусов текстов, использование которых необходимо в процессе обучения и при проведении научных исследований. В третьей главе раскрываются возможности инструмента PRD для анализа ритма прозы. В четвертой представлен комплексный анализ ритма текста с опорой на инструмент, а также представлены возможности проведения сравнительного анализа оригинала и перевода в русском и английском языках. В пятой главе описывается механизм проведения статистических экспериментов с использованием инструмента PRD.

Каждый раздел сопровождается заданиями и вопросами по темам.

УДК 004
ББК 32

ISBN 978-5-91730-998-9

© Бойчук Е.И., Лагутина Н.С., Лагутина К.В., Воронцова И.А., Шляхтина Е.В., Мишенькина Е.В., Беляева О.В.

СОДЕРЖАНИЕ

Глава I. ИНСТРУМЕНТЫ И ПРИЛОЖЕНИЯ ДЛЯ РАБОТЫ С РАЗЛИЧНЫМИ ТИПАМИ ТЕКСТОВ	4
§1. Онлайн ресурсы для семантического анализа текста	4
§2. Частотный анализ текста, оценка читабельности и фоносемантический анализ.....	8
§3. Рифмовники	12
§4. Описание инструмента #LancsBox 5.0	16
Глава II. ИСПОЛЬЗОВАНИЕ КОРПУСОВ ТЕКСТОВ В УЧЕБНЫХ И НАУЧНЫХ ЦЕЛЯХ	26
§1. Национальные корпуса языка	26
§2. Обзор корпусов текстов, аудио- и видеоматериалов для учебных целей.....	28
§3. Обзор корпусов текстов для научных целей.....	33
Глава III. ИНСТРУМЕНТ ДЛЯ АНАЛИЗА РИТМА ПРОЗЫ PROSE RHYTHM DETECTOR (PRD)	35
§1. Предварительная обработка текста	35
§2. Ритмические средства для автоматизированного анализа	51
§3. Алгоритмы поиска ритмических средств	54
Глава IV. ЛИНГВИСТИЧЕСКИЕ ВОЗМОЖНОСТИ ИНСТРУМЕНТА PRD	56
§1. Анализ ритма художественного текста.....	56
§2. Анализ ритмической структуры текста и его перевода.....	61
Глава V. ПРОВЕДЕНИЕ СТАТИСТИЧЕСКИХ ЭКСПЕРИМЕНТОВ С ИСПОЛЬЗОВАНИЕМ PRD	72
§1. Составление корпуса текстов.....	72
§2. Оценка качества разметки	75
§3. Статистические характеристики.....	77
§4. Визуализация результатов.....	79
БИБЛИОГРАФИЯ	86

Учебное пособие

**Бойчук Е.И., Лагутина Н.С., Лагутина К.В., Воронцова И.А.,
Шляхтина Е.В., Мишенькина Е.В., Беляева О.В.**

АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ ТЕКСТОВЫХ ХАРАКТЕРИСТИК

Компьютерная верстка С.С. Ламан

Подписано в печать 28.11.2020 г.
Объем 11 п.л. Тираж 100 экз. Заказ № 5415.

Отпечатано с предоставленных оригинал-макетов
в типографии «Канцлер»
150008, г. Ярославль, ул. Полушкина роща, д. 16, стр. 66а.
Тел.: 8-4852-58-76-33, 8-4852-58-76-39
E-mail: kanclel2007@yandex.ru