

Научная группа Компьютерная лингвистика и автоматический анализ текстов

Внесена в Реестр Научных школ и Научных групп ЯрГУ 22.02.2023 г., Приказ № 172.

Руководитель и основатель	Парамонов Илья Вячеславович - доцент, кандидат физико-математических наук, доцент кафедры компьютерных сетей ЯрГУ, ORCID , AuthorID
Коды ГРНТИ	16.31.21 Автоматическая обработка текста. Автоматический перевод. Автоматическое распознавание речи 28.23.02 Общие проблемы искусственного интеллекта
Основные результаты	Разработаны алгоритмы для определения основных параметров ритма текста на лексико-грамматическом уровне (ассонанс, аллитерация, анафора, эпифора, симплока, анадиплосис, эпаналепсис, диакопа, эпизевксис, хиазм, полисиндетон, последовательность вопросительных предложений, последовательность восклицательных предложений, апозиопеза) и на фоностилистическом уровне (ассонанс и аллитерация). Для данных средств определены условия их употребления для английского, французского, испанского и русского языков, сформированы списки исключений и специфического употребления тех или иных форм. Разработан программный инструмент ProseRhythmDetector, представляющий собой первый в мире программный инструмент, позволяющий осуществить анализ ритма прозы с точки зрения фоностилистического и лексико-грамматического аспектов. Экспериментально показано, что ритмические характеристики как самостоятельный набор параметров текстов, эффективно применим для решения задачи верификации авторов художественных произведений. Разработаны алгоритмы автоматического построения тезаурусов предметных областей на основе методов выделения ключевых слов для английского и русского языков. Результаты исследований по созданию тезаурусов были апробированы на задаче анализа тональности текстов и задаче классификации текстов по предметным областям. В обоих случаях было достигнуто повышение качества решения основной задачи по сравнению со случаями, когда тезаурусы не использовались. Выполнена адаптация метода анализа тональности русскоязычных текстов на основе синтаксической структуры предложений, а также проведен сравнительный анализ эффективности различных нейросетевых классификаторов для задачи классификации русскоязычных классов по тональности на четыре класса (положительный, отрицательный, нейтральный и смешанный).
Члены научного коллектива	Лагутина Надежда Станиславовна ORCID , AuthorID Лагутина Ксения Владимировна ORCID , AuthorID Васильев Андрей Михайлович ORCID , AuthorID Мельникова Оксана Андреевна ORCID , AuthorID Полетаев Анатолий Юрьевич ORCID , AuthorID Костерин Максим Алексеевич ORCID , AuthorID Зафиевский Даниил Дмитриевич ORCID
Контакты	i.paramonov@uniyar.ac.ru