

ОТЗЫВ

члена диссертационного совета

БУРЕ Владимира Мансуровича

на диссертацию Тарасова Никиты Андреевича на тему «Гибридные нейросетевые методы анализа понятности текстов юридических документов на русском языке», представленную на соискание ученой степени кандидата технических наук по научной специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Общая информация о диссертации

Диссертация Тарасова Н. А. состоит из введения, пяти глав, заключения, списка литературы, включающего 156 наименований использованных источников. Общий объем текста на русском языке составляет 124 страницы, включая 21 рисунок и 14 таблиц.

Материалы диссертации опубликованы в 9 печатных работах, из них 4 статьи в изданиях Scopus и WoS, 5 в тезисах докладов, получены 3 свидетельства о государственной регистрации программ для ЭВМ.

Актуальность темы исследования

Автоматизация обработки текстовых данных позволяет эффективно решать широкий спектр задач анализа юридических документов. Использование современных нейросетевых методов позволяет обрабатывать большие объемы правовых текстов, решая целый ряд задач, возникающих в результате правового процесса. Интеллектуальные методы анализа текстов дают возможность классифицировать юридические тексты по различным языковым признакам, в том числе сложности восприятия. Улучшение доступности восприятия официальных юридических текстов помогает перейти к более прозрачному взаимодействию юристов с населением в т.ч. в вопросах законодательства.

Достоверность выводов и результатов исследования

Достоверность выполненных автором исследований подтверждена результатами вычислительных экспериментов и адекватностью использованных в работе разработанных автором методов и алгоритмов.

Научная новизна исследований и полученных результатов

В исследовании была представлена методология автоматизированной классификации юридических текстов по сложности, сочетающая классические методы машинного обучения с нейросетевыми языковыми моделями. Была разработана система, способная ассоциировать определенные типы языковых конструкций с параметрами доступности текста, что позволило существенно улучшить точность оценки сложности юридических документов.

Теоретическая и практическая значимость работы

Результаты исследования имеют как теоретическое, так и практическое значение. Создан набор методологических и программных инструментов, направленных на оценку сложности и доступности юридических документов. Путём применения статистики и машинного обучения были определены характеристики, которые эффективно моделируют

сложность текстов правового характера. Программный комплекс, основанный на современных языковых моделях, позволяет оценивать сложность таких текстов.

Анализ содержания работы

Введение включает следующие элементы: актуальность темы исследования; цель и задачи; научная новизна; теоретическая и практическая значимость работы; апробация результатов; информация о публикациях; структуре и объеме работы; основные научные результаты; положения, выносимые на защиту. В первой главе рассматривается использование статистических методов для оценки частоты юридических терминов. Вторая глава предоставляет обзор признаков, определяющих сложность юридических документов, оценивает эффективность их использования в задаче классификации по сложности. Третья глава описывает гибридный подход, объединяющий языковые характеристики с нейросетевыми языковыми моделями. В четвертой главе приводится анализ сложности документов разной стилистики с помощью созданной модели, а так же описывается процесс сбора и обработки данных. В пятой главе продемонстрирован пример применения методов в задаче анализа ответов налоговых служб.

Замечания

Имеется ряд замечаний:

1. В разделе 3.6 главы “Гибридная модель оценки сложности: разработка и применение для российских юридических текстов” не хватает более подробного описания методологии выбора модели, как следствие, затруднен анализ таблицы со значениями метрик качества.
2. В разделе 4.4.1 главы “Языковая сложность русских юридических подстилей и жанров” не хватает пояснений по наиболее качественным метрикам - необходима расшифровка аббревиатур, приведение формул.
3. В главе “Метрики сложности российских юридических текстов: выбор, использование, первоначальная оценка эффективности” отсутствует математическое представление формул читаемости.

Заключение

Диссертация представляет собой целостную и законченную научно-исследовательскую работу на актуальную тему, выполненную автором самостоятельно на хорошем научно-техническом уровне. Основные выводы по результатам исследований достоверны и обоснованы. Опубликованные научные работы подробно отражают основное содержание диссертации и полученные выводы.

Диссертационная работа Тарасова Никиты Андреевича «Гибридные нейросетевые методы анализа понятности текстов юридических документов на русском языке» по теоретическому уровню и практической значимости соответствуют требованиям «Положения о присуждении ученых степеней», предъявляемым к кандидатским диссертациям, а её автор, Тарасов Никита Андреевич, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Диссертация Тарасова Никиты Андреевича на тему: «Гибридные нейросетевые методы анализа понятности текстов юридических документов на русском языке» соответствует основным требованиям, установленным Приказом от 19.11.2021 № 11181/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», соискатель Тарасов Никита Андреевич заслуживает присуждения ученой степени кандидата технических наук по научной специальности 2.3.1. Системный анализ, управление и обработка информации, статистика. Нарушения пунктов 9 и 11 указанного Порядка в диссертации не обнаружены.

Член диссертационного совета
Доктор технических наук,
Профессор кафедры МТИиСР СПбГУ
19804, Санкт-Петербург, Университетский пр., 35,
Тел. +7(911)957-58-60, v.bure@spbu.ru
Буре Владимир Мансурович

Дата

25.11.2024



Подпись