



Міністэрства адукацыі Рэспублікі Беларусь

УСТАНОВА АДУКАЦЫІ
«ПОЛАЦКІ
ДЗЯРЖАЎНЫ ЎНІВЕРСІТЭТ»

вул. Блахіна, 29

211446, г. Наваполацк Віцебскай вобл.

Тэл.: (8 0214) 50 57 00; факс: (8 0214) 59 95 36

E-mail: post@psu.by; вэб-рэсурс: www.psu.by

р/р ВУ68АКВВ36049000028952100000 (бюджэт)

ВУ32АКВВ36329000028962100000 (пазабюджэт)

у ЦБП № 214 ААТ «ААБ Беларусбанк»

г. Наваполацк, вул. Дружбы, 4, БИК АКВВВУ2Х

АКПА 02071694, УНП 300220696

Министерство образования Республики Беларусь

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«ПОЛОЦКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ул. Блохина, 29

211446, г. Новополоцк Витебской обл.

Тел.: (8 0214) 50 57 00; факс: (8 0214) 59 95 36

E-mail: post@psu.by; веб-ресурс: www.psu.by

р/с ВУ68АКВВ36049000028952100000 (бюджет)

ВУ32АКВВ36329000028962100000 (внебюджет)

в ЦБУ № 214 ОАО «АСБ Беларусбанк»

г. Новополоцк, ул. Дружбы, 4, БИК АКВВВУ2Х

ОКПО 02071694, УНП 300220696

04.12.2024 № 04-12-80

На № _____ ад _____ **ОТЗЫВ**

Члена диссертационного совета, доктора технических наук Богуш Рихарда Петровича 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

Существует достаточно большое количество групп юридических документов, которые используются для решения огромного множества задач. Известно, что такие документы характеризуются различной сложностью, а пользователи обладают разным уровнем знаний в данной области, поэтому эффективность взаимодействия между пользователями очень низкая. Существующие и развивающиеся методы анализа текстовых данных на основе нейронных сетей и машинного обучения предоставляют возможность для автоматизации обработки юридической документации. Поэтому тема диссертационной работы и заявленная цель на разработку методологических и инструментальных средств интеллектуальной обработки юридических текстов и алгоритмического обеспечения процесса определения доступности их восприятия являются актуальными в настоящее время.

Структура и содержание диссертации достаточно полно соответствуют поставленной цели и раскрывают полученные результаты. В первой главе содержится информация по существующим методам определения лексической сложности текстов на русском языке. Вторая глава посвящена проблеме выбора метрик понятности юридических документов. Третья глава включает разработанный гибридный метод оценки сложности, основанный на совместном применении лингвистических характеристик и большой нейросетевой языковой модели. В четвертой главе оценивается языковая сложность русских юридических документов различных стилей и типов из различных источников. В пятой главе анализируется уровень доступности восприятия письменных ответов налоговых органов на обращения физических лиц и организаций с применением разработанной автором гибридной семантической модели. В заключении отражены основные результаты, список литературы содержит 156 наименований источников и является вполне достаточным по теме диссертации.

К основным научным результатам следует отнести: методологию оценки понятности юридических документов разных стилей и типов на русском языке; гибридный метод оценки сложности русских юридических документов на основе лексических характеристик и больших языковых нейросетевых моделей; комплекс программ для информационного обеспечения научно-исследовательской и опытной работы с русскоязычными текстовыми данными правовых процессов, включающий компоненты сбора, интеллектуального анализа и визуализации результатов. Необходимо отметить, что исследования автора диссертации могут способствовать повышению эффективности работы с юридическими документами как профильных специалистов, так и граждан при их обращении в государственные органы.

Достоверность выполненных автором исследований подтверждена результатами вычислительных экспериментов. Материалы диссертации опубликованы в 9 печатных работах, из них 4 статьи в изданиях Scopus и WoS, 5 тезисов докладов. Автор разработал программное обеспечение и получил 3 свидетельства о государственной регистрации программ для персональных компьютеров. Считаю, что данный перечень работ является достаточным для представления и апробации материалов кандидатской диссертации.

К замечаниям следует отнести следующие:

1. Пункт 1.4 не в полной мере раскрывает методы моделирования общезыковых частот и диапазонов частот, так как он не содержит четкого анализа таких методов с выявлением их достоинств и недостатков.

2. Следовало бы использовать строгие научные и технические термины вместо следующих: «низкочастотных хвостов» (стр.18), «изопрённым» (стр.24), «более трезво» (стр.26) и т.п.

В работе не определен термин «веб-корпус».

3. В выводах по первой главе отмечено «Целью было не само сравнение, а разработка методологии создания сводного списка частот и моделирования общезыковой частотности». При данном утверждении возникает вопрос, в каком пункте гл.1 уже сформулирована разработанная методология?

4. В п.2.4.2 не сказано, в каком соотношении выполнено разбиение исходной выборки на обучающую и тестовую, а также почему не выделен такой тип выборки, как валидационный набор. При его наличии объективность оценки работы модели была бы выше. Аналогичное замечание справедливо и для описания исследований в других пунктах диссертации.

5. Из выводов по второй главе следует, что тестирование проводилось на наборах данных, существенно отличающихся от целевых. Нет пояснения, почему так было сделано и почему не использовались изначально выборки, содержащие необходимые данные.

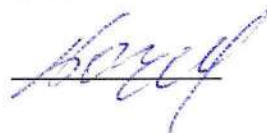
6. В п.3.4 представлен перечень типов лингвистических характеристик, первая из которых в этом списке «Основные признаки». При дальнейшем описании наименование п.3.4.1 Базовые метрики. Отсюда следует понимать, что термин «признаки» соответствует термину «метрики» или допущена некорректность при изложении материала диссертации?

Отмеченные замечания не снижают высокую научную значимость полученных автором результатов и незначительно влияют на положительное впечатление о работе.

Таким образом, проведенный мною анализ данной работы показал, что диссертация представляет собой целостную и законченную научно-исследовательскую работу на актуальную тему, выполненную автором самостоятельно на достаточном научно-техническом уровне. Основные выводы по результатам исследований достоверны и обоснованы. Опубликованные научные работы в полной мере отражают содержание диссертации и положения, выносимые на защиту.

Диссертационная работа Тарасова Никиты Андреевича «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, доцент,
заведующий кафедрой вычислительных систем и сетей
Полоцкого государственного университета
имени Евфросинии Полоцкой



Р.П. Богуш

Подпись Богуша
Специалист

