

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Животовой Алены Анатольевны

«Математическая модель, алгоритмы и программный комплекс для повышения качества машинного перевода узкоспециальных технических текстов на английский язык», представленную к защите на соискание ученой степени

кандидата технических наук по специальности

1.2.2 – «Математическое моделирование, численные методы и комплексы программ»

1. Структура и объем работы

Рецензируемая диссертационная работа представлена:

- диссертацией, состоящей из введения, четырех глав, заключения, библиографического списка и двух приложений. Диссертация содержит 125 страниц, 22 рисунка и 13 таблиц. Список литературы включает 124 наименования;
- авторефератом на 24 страницах со списком опубликованных работ автора.

2. Актуальность темы диссертационного исследования

Диссертационная работа Животовой А.А. посвящена разработке математических моделей, алгоритмов и программного комплекса, которые позволяют оценивать сложность задач перевода и оптимизировать структуру исходных текстов с целью повышения качества машинного перевода узкоспециальных технических текстов с русского на английский язык. Актуальность исследования определяется несколькими ключевыми факторами, связанными с современными вызовами в области машинного перевода, особенно в контексте узкоспециальных технических текстов. В условиях роста объемов создаваемой информации и необходимости ее локализации в реальном времени, качество перевода становится критически важным для эффективной коммуникации во многих специализированных технических областях, где точность и своевременность перевода могут существенно повлиять на результаты ключевых проектов.

Современные системы машинного перевода, несмотря на значительные достижения в области нейросетевых технологий и больших языковых моделей, продолжают сталкиваться с проблемами, связанными с качеством перевода, особенно в узкоспециальных областях, где сбор достаточного объема двуязычных тренировочных данных затруднен. Это часто обусловлено тем, что используемые при переводе соответствия редко используются в других областях и являются специфическими как раз для узкоспециализированной области. Как следствие, требуется разработка новых методов предредактирования текстов, которые могут повысить качество перевода и снизить зависимость от человеческого фактора. Представленное исследование направлено на решение этой проблемы и его актуальность не вызывает сомнений.

3. Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации, их достоверность и новизна

Анализ содержания представленной работы и автореферата производит положительное впечатление о проведенном исследовании, которое включает все составляющие, определенные специальностью 1.2.2. – «Математическое моделирование, численные методы и комплексы программ». Автореферат диссертации полно отражает основные положения и выводы, к которым пришел автор в ходе работы.

Основная идея исследования заключается в автоматизации предредактирования текстов с целью оптимизации их структуры, что позволит системам машинного перевода более эффективно выполнять перевод и снижать количество стилистических ошибок. В

работе предложены новые методики и алгоритмы, включая вероятностную оценку сложности задач перевода, использование обратного перевода для сбора тренировочных данных.

Содержание работы охватывает несколько ключевых аспектов, начиная с анализа существующих систем машинного перевода и заканчивая практической реализацией разработанных алгоритмов в программном комплексе. Перечень глав и последовательность их представления в диссертации соответствуют выполнению поставленных научных задач, решение которых раскрывает суть изучаемой проблемы.

В первой главе рассматривается предредактирование как важное направление развития систем машинного перевода. Автор анализирует эволюцию технологий перевода, включая интерактивный перевод и методы предредактирования, направленные на улучшение качества перевода.

Вторая глава посвящена разработке математической модели оптимизационного предредактирования. Представлена обобщенная модель процесса перевода, включающая этапы обучения, постановки задачи, оценки сложности задачи перевода, предредактирования, перевода, оценки качества и постредактирования. Также разработан метод оценки сложности задачи перевода, что позволяет прогнозировать риски, связанные с качеством перевода в зависимости от компетенции переводчика и характеристик исходного текста.

Третья глава описывает методику реализации оптимизационного предредактирования. В ней рассматриваются алгоритмы обучения моделей, предназначенных для оценки сложности задач перевода и оптимизации исходных текстов.

Четвертая глава посвящена созданию программного комплекса, реализующего предложенные алгоритмы. Описывается архитектура комплекса и полученные в ходе реализации описанных методов и алгоритмов массивы данных: база данных показателей структурного анализа предложений технических русскоязычных текстов, корпус параллельных двуязычных текстов нефтегазовой тематики для тренировки языковых моделей в задачах перефразирования узкоспециальных технических русскоязычных текстов и повышения качества их перевода на английский язык. Приводятся результаты тестирования программного корпуса на реальных данных, описан процесс внедрения и тестирования в рамках деятельности компании, предоставляющей услуги перевода технической документации нефтегазовых проектов.

В **заключении** диссертационной работы даны формулировки основных результатов работы.

Список использованной литературы содержит 124 работы, в том числе подготовленные автором при выполнении диссертационного исследования.

В **приложении** представлены охранные документы на результаты интеллектуальной собственности и акт внедрения результатов работы.

Научная новизна представленных в работе положений и результатов определяется следующим образом:

1. Предложена новая методика для повышения качества машинного перевода текстов с русского языка на английский язык, отличающаяся от существующих применением обратного перевода для сбора тренировочных данных и оптимизационного предредактирования на основе вероятностной оценки сложности задачи перевода.

2. Впервые предложена методика оценки сложности переводческой задачи для переводчика на основе его компетенции и специализации и параметров исходного текста, которая позволяет прогнозировать риски некачественного и/или несвоевременного решения задачи перевода.

3. Предложен новый алгоритм, позволяющий расширить область применения оптимизационного метода градиентного спуска путём использования элементов нечеткой логики в выражении функции правдоподобия через функцию принадлежности

полученного текста низкой сложности задачи перевода для выбранной системы машинного перевода.

4. Предложен новый алгоритм, позволяющий расширить область применения метода наименьших квадратов для поиска весов значимости параметров исходного текста для вероятностной оценки ожидаемого качества его перевода на целевой язык.

5. Предложена новая архитектура и реализован программный комплекс для повышения качества машинного перевода текстов с русского языка на английский язык, отличающийся от существующих применением ансамбля моделей для оптимизационного предредактирования на основе вероятностной оценки сложности задачи перевода с целью повышения качества машинного перевода текстов с русского языка на английский язык.

Обоснованность и достоверность результатов исследования обеспечивается применением апробированных математических методов, таких как теория множеств, численные методы оптимизации, включая метод наименьших квадратов и метод градиентного спуска, а также статистические методы, в частности метод максимального правдоподобия. Кроме того, достоверность результатов подтверждается применением современных программных комплексов анализа данных, что обеспечивает детальную верификацию полученных результатов. Экспериментальная апробация разработанных алгоритмов на корпусе узкоспециальных текстов также служит дополнительным фактором, подтверждающим достоверность выводов исследования.

Результаты, изложенные в диссертации, отражают личный вклад соискателя. Квалификационный уровень и научную ценность полученных результатов подтверждает достаточное количество работ, опубликованных в рецензируемых научных изданиях. Общий список работ включает 22 научных работы, из них 3 статьи в журналах, рекомендованных ВАК России, 3 – в изданиях, индексируемых в базе Scopus, 6 публикаций в трудах конференций. Автором получено 10 свидетельств о государственной регистрации программ и баз данных для ЭВМ.

Таким образом, диссертация представляет собой комплексное исследование, в котором обоснованы новые подходы к предредактированию текстов, разработаны математические модели и алгоритмы, расширена область применения численных методов оптимизации, а также создан программный комплекс, способствующий повышению качества машинного перевода и производительности пост-редакторов перевода. Обоснованность и достоверность результатов и выводов диссертационной работы не вызывает сомнений.

4. Практическая и теоретическая значимость работы

Диссертация вносит вклад в теорию и практику машинного перевода, предлагая оригинальные подходы и решения, которые могут быть использованы как в научных исследованиях, так и в практической деятельности переводческих компаний.

Теоретическая значимость проведенного исследования заключается в его вкладе в развитие методов предредактирования узкоспециальных текстов для повышения качества их машинного перевода с русского языка на английский. Работа адресует недостаточную проработанность данной темы в научной литературе и, основываясь на теории множеств и методах оптимизации, предлагает математические модели и методы, которые позволяют оценивать сложность задач перевода и автоматизировать подготовку русскоязычных текстов к переводу. Это расширяет теоретическую базу в области машинного перевода и обработки естественного языка и создает основу для дальнейших исследований в данной области. Также, разработанные в ходе исследования базы данных параллельных корпусов двуязычных текстов могут быть использованы в исследовательских целях в смежных областях, связанных с обработкой и анализом русскоязычных текстов и их перевода на английский язык.

Разработанные алгоритмы и программный комплекс, основанные на вероятностной оценке сложности задач перевода и методах оптимизационного предредактирования, могут быть интегрированы в существующие системы управления и автоматизации переводческой деятельности, что позволит повысить качество перевода узкоспециальных текстов, оптимизировать затраты и снизить зависимость от человеческого фактора. Результаты исследования уже внедрены в практическую деятельность предприятия лингвистической отрасли, что подтверждает их реальную применимость и эффективность. Акт внедрения и свидетельства о регистрации программ ЭВМ и баз данных демонстрируют, что разработанные алгоритмы и программные комплексы успешно функционируют в условиях реального использования, что подчеркивает их практическую значимость.

5. Замечания и вопросы по содержанию диссертационной работы

Необходимо отметить ряд вопросов по содержанию диссертационной работы:

1. В рамках описания обобщенной модели процесса перевода в разделе 2.1.1 на стр. 47-48 автор определяет формализованные понятия специализации и компетенции переводчика. Далее в тексте диссертации недостаточно полно раскрыт вопрос о том, как данные понятия на практике оцениваются для систем машинного перевода.

2. Для решения задачи оптимизации в контексте предредактирования текстов используется метод градиентного спуска (стр. 67), однако не приведены аргументы, объясняющие этот выбор.

3. В главе 4 подробно описана архитектура разработанного программного комплекса, однако не описана методология интеграции комплекса в контур автоматизации переводческих компаний.

4. В диссертации ключевым понятием и контрольным параметром является качество перевода, однако данное понятие можно трактовать по-разному, оно носит субъективный характер. В области машинного разработаны методы оценки качества автоматизированного перевода, но они обладают разным уровнем точности и результат их оценки может не совпадать с оценкой пользователя перевода. Каким образом достигается объективность и точность в вопросе оценки качества машинного перевода?

6. Общее заключение

Вышеперечисленные замечания не меняют общего положительного впечатления о диссертационной работе.

Считаю, что диссертация Животовой Алены Анатольевны обладает новизной и практической значимостью, представляет решение актуальных междисциплинарных задач, объединенных общим подходом, имеющих значение для дальнейшей разработки проблематики повышения качества машинного перевода русскоязычных текстов на английский язык с использованием методов автоматического предредактирования. Полученные автором научные результаты исследования прошли апробацию на научных конференциях, а также отражены в достаточном количестве научных публикаций.

Все вышеизложенное позволяет заключить, что диссертационная работа «Математическая модель, алгоритмы и программный комплекс для повышения качества машинного перевода узкоспециальных технических текстов на английский язык» является завершенной самостоятельной научно-квалификационной работой.

Тематика и результаты выполненного исследования соответствуют пп. 2, 3, 4, 8 паспорта специальности 1.2.2 – «Математическое моделирование, численные методы и комплексы программ».

Представленная работа соответствует требованиям, установленным пп. 9-11 Положения о присуждении ученых степеней, утвержденного постановлением

Правительства РФ, а ее автор, Животова Алена Анатольевна, заслуживает присуждения ученой степени кандидата технических наук по специальности 1.2.2 – «Математическое моделирование, численные методы и комплексы программ».

Официальный оппонент,
доктор технических наук по
специальности 05.13.17,
профессор, заместитель
директора по науке Института
математики и компьютерных
технологий ФГАОУ ВО
«Дальневосточный федеральный
университет»

Артемьева Ирина Леонидовна

«13» 09 2024 г.

Почтовый адрес:
690922, г. Владивосток,
о. Русский, п. Аякс, д.10
Тел: +79147015642
Email: artemeva.il@dvfu.ru

Я, Артемьева Ирина Леонидовна, даю согласие на включение моих персональных данных в документы, связанные с работой диссертационного совета, и их обработку.

Личную подпись официального оппонента Артемьевой И. Л. заверяю

