

Отзыв о квалификационной работе бакалавра Лыщика Андрея Игоревича.

В рамках квалификационной работы Андрей исследовал существующие методы автоматического определения эмоциональной окраски текстов, подготовил набор данных для анализа применимости этих методов к текстам на русском языке, а также провел сравнительный анализ качества работы различных методов.

Выбранная автором тема, безусловно, интересна как с практической, так и с теоретической точки зрения. Эмоциональная окраска текстов, как правило рецензий и отзывов, является важным источником информации для решения задачи продвижения товаров и брендов, социологических исследований. При этом на данный момент задачу нельзя считать окончательно решенной, в особенности для русского языка.

В работе Андрея приводится обширный анализ существующих работ, посвященных анализу эмоциональной окраски текстов на английском языке с помощью разнообразных техник машинного обучения. Из рассмотренных работ выбраны наиболее перспективные, для их сравнительного анализа подготовлен набор данных на базе рецензий с сайта «Кинопоиск». Анализ качества проведен с помощью метода перекрестной валидации и анализом статистической значимости результатов. Кроме того, исследована возможность построения ансамбля алгоритмов для улучшения качества прогноза.

Среди недостатков работы можно в первую очередь отметить отсутствие в сравнении методов, основанных на многослойных нейронных сетях, которые в последнее время активно развиваются в области анализа текстов.

В процессе работы Андрей показал себя как грамотный и квалифицированный исследователь с высоким уровнем ответственности и самоорганизации.

Рекомендую оценить квалификационную работу Лыщика Андрея на «отлично» и допустить его к защите.

научный руководитель,
кандидат физико-математических наук
Бугайченко Дмитрий Юрьевич



Бугайченко Д.И.

Отзыв о квалификационной работе бакалавра Ананьева Дениса Владиславович.

В рамках квалификационной работы Денис исследовал существующие инструменты автоматического определения языка, на котором написан текст, проанализировал работу этих инструментов, подготовил корпус текстов на языках стран СНГ, натренировал языковые профили на этом корпусе и проанализировал качество работы, как на исходном корпусе, так и на корпусе реальных пользовательских текстов социальной сети.

Выбранная автором тема, безусловно, интересна как с практической, так и с теоретической точки зрения. Для латинских языков тема исследована достаточно широко, но для языков стран СНГ (узбекский, таджикский, и т. д.) готовых инструментов на данный момент нет. При этом определение языка часто является важным подготовительным шагом для решения других задач анализа текстов: определения семантики и эмоциональной окраски, индексации, синтаксическому разбору.

В работе Денис отталкивается от двух популярных библиотек определения языка, одна из которых широко распространена в индустрии. В контексте поддержки языков стран СНГ «из коробки» поддерживается только русский, украинский и белорусский. Для того, чтобы иметь возможность определять другие языки стран СНГ на базе текстов википедии соответствующих языковых версий был подготовлен размеченный корпус и натренированы профили этих языков. После чего проанализировано качество работы стандартных и новых профилей на этом же корпусе, а также анализ качества работы на данных с постами пользователей социальной сети «Одноклассники».

На данный момент натренированные Денисом профили языков проходят пилотное внедрение в компании Mail.ru.

Среди недостатков работы следует отметить достаточно поверхностный анализ качества работы на пользовательских данных, а также отсутствие анализа влияния дополнительных методов очистки и нормализации текстов на качество определения языка.

В процессе работы Денис показал себя, с одной стороны, как человек, способный решить поставленную задачу, но, с другой стороны, продемонстрировал высокую степень нестабильности и неорганизованности.

Рекомендую оценить квалификационную работу Ананьева Дениса на «хорошо» и допустить его к защите.

научный руководитель,
кандидат физико-математических наук
Бугайченко Дмитрий Юрьевич

Бугайченко Д.Ю.