

РЕЦЕНЗИЯ

на выпускную квалификационную работу обучающегося СПбГУ

Андреевой Анны Леонидовны

по теме «Автоматическое выделение терминов, образованных от имен собственных»

Квалификационное исследование А.Л. Андреевой посвящено актуальной проблеме в рамках такого популярного направления научных изысканий, как методики автоматического извлечения терминов из текстов на естественном языке – разработке и реализации алгоритма извлечения терминов, образованных от имен собственных, или терминов-эпонимов. В то время как вопрос о статусе и структуре терминов-эпонимов поднимался в рамках терминоведения, до настоящего времени не было предложено формальных методов, позволяющих распознавать их в специальном тексте и извлекать из него.

Указанную цель исследования А.Л. Андреева достигает путем последовательного выполнения 7 задач как теоретического, так и инженерно-практического характера; все они успешно решены в ходе исследования, что отражено в тексте ВКР. Работа объемом 53 страницы состоит из введения, двух глав, заключения, списка литературы и четырех приложений. Первая глава представляет собой теоретический обзор концепций, связанных с понятием термина, включая описанные в литературе подходы к определению термина, классификации терминов, особенности их функционирования, формальную структуру и способы образования. Далее логичным образом следует переход к понятию термина-эпонима и к употреблению таких терминов в специальной литературе. В целом эта часть работы безусловно свидетельствует о глубокой эрудиции А.Л. Андреевой в области терминоведения; подбор литературы для обзора включает в себя как классические работы 50–80-х гг. XX в., так и издания последних лет. Можно

без колебаний констатировать, что уровень рассмотрения теоретических вопросов автором соответствовал бы и работе более высокого статуса, чем выпускная квалификационная работа бакалавра.

Во второй главе А.Л. Андреева излагает методику извлечения терминов-эпонимов из научного текста. Разработанный автором алгоритм сформулирован теоретически и затем реализован на языке программирования Python, после чего приведен анализ полученных результатов. Материал исследования включает терминологические конструкции трех различных типов, потребовавшие различных подходов при обработке. А.Л. Андреева дает обзор статистических и лингвистических методов, применяемых при выделении терминов, и убедительно обосновывает выбор в пользу одного из них либо их комбинации для каждой конкретной конструкции. В результате применения алгоритма доля правильно выделенных терминов-эпонимов ни для одной из групп не составляет менее 64%. Констатируя работоспособность алгоритма, А.Л. Андреева, однако, признает необходимость его доработки с целью повышения точности. Приведенный анализ ошибок программы и предлагаемый комплекс мер по улучшению работы алгоритма вполне убедителен, и в целом описанное в главе исследование, равно как и всю работу, следует признать проведенным на весьма высоком уровне.

Обращаясь к формальной стороне рецензируемого текста, следует отметить хороший, ясный язык, которым написана работа, и ее безукоризненный научный стиль. Структура ВКР также является четкой и логичной и способствует пониманию содержания. На этом фоне несущественными недочетами представляются выравнивание текста по левому краю, а не по ширине; отсутствие названий двух основных глав при наличии названий у каждого из разделов внутри каждой главы; отсутствие точек после некоторых элементов списков во введении, в тексте глав и в списке литературы, а также отдельные обнаруженные опечатки. Однако чтение работы все же оставило без ответа некоторые вопросы, которые автор рецензии хотел бы адресовать автору ВКР.

• На с. 36 (раздел 2.4 «Анализ полученных результатов») приводится таблица результатов работы алгоритма для каждого из трех типов исследуемых терминологических конструкций в числовом выражении. Стоит отметить, что приведенные числа соответствуют такому параметру, как *полнота*, указывая процент извлеченных терминологических словосочетаний от общего их числа по корпусу. Однако для двух типов конструкций (прилагательное + существительное, существительное + существительное в родительном падеже) приводятся также данные о словосочетаниях, не являющихся терминологическими, но выделенных алгоритмом (*точность*), правда, без цифр. Рассматривает ли автор достижение полноты при автоматическом выделении терминологических словосочетаний как более важную проблему и направлены ли меры по улучшению алгоритма, приведенные в конце раздела, только на увеличение полноты или на увеличение точности тоже?

• На с. 4 введения и далее на с. 31 раздела 2.2 «Сбор материалов для исследования» автор описывает процесс создания словаря, на материале которого далее был составлен корпус, послуживший входными данными для программы. При этом указывается, что для каждого из 100 терминов из словаря было отобрано ровно по 3 случая употребления; при этом данная цифра в тексте работы никак не обосновывается. Несмотря на то, что данный вопрос представляется в контексте исследования скорее второстепенным, нам все же хотелось бы прояснить его.

• В разделе 1.7 «Употребление терминов-эпонимов в специальной литературе» автор указывает, что сфера употребления терминов, образованных от имен собственных, не ограничивается точными и естественными науками, но также включает в себя такие дисциплины, как лингвистика, юриспруденция, история, экономика и др. Однако же, термины, отобранные для практической части работы, взяты исключительно из текстов по естественным наукам – медицине, математики, химии и физике. В связи с этим встает вопрос об универсальности разработанного алгоритма. Может ли

он быть с минимальными изменениями применен к текстам по другим специальностям? Или же алгоритм мыслится в первую очередь как способ извлечения *естественнонаучных* терминов-эпонимов?

Вне зависимости от ответов, которые могут быть получены на поставленные вопросы, выпускную квалификационную работу А.Л. Андреевой следует признать полноценным научным исследованием, проведенным на достойном уровне и заслуживающим высокой оценки.

«12» июня 2018 г.

Подпись

Филиппов А.К.
ФИО