

ОТЗЫВ

официального оппонента на диссертацию Аветисяна Карена Ишхановича
«Метод обнаружения межъязыковых заимствований в текстах»,
представленную на соискание учёной степени кандидата технических наук
по специальности 2.3.5 «Математическое и программное обеспечение
вычислительных систем, комплексов и компьютерных сетей»

Актуальность темы. Одной из важнейших задач в рамках автоматической обработки текстов является обнаружение текстовых заимствований. Влияние достижений в этой области на научно-образовательную сферу сложно переоценить: множество вузов и научных организаций во всем мире используют системы проверки на плагиат при оценке различных видов научных работ, включая квалификационные. Однако в последнее время в связи с развитием систем машинного перевода стала актуальной проблема выявления межъязыковых заимствований, когда в работе на одном языке используются фрагменты работ, исходно написанных на другом языке. Особенно острой эта проблема является для малоресурсных языков, например, армянского, для которых отсутствуют или слабо доступны качественные системы машинного перевода. Поэтому диссертационная работа К. И. Аветисяна, посвященная разработке метода и программных средств обнаружения межъязыковых заимствований, применимых в том числе к малоресурсным языкам, является весьма своевременной и востребованной.

Научная новизна. В диссертационной работе получены следующие наиболее важные научные результаты:

1. Разработан метод обнаружения межъязыковых заимствований, применимый к большому количеству языков и, в частности, к малоресурсным языкам, для которых системы машинного перевода и другие инструменты автоматической обработки текстов либо отсутствуют, либо имеют недостаточно высокое качество.

2. Предложен метод генерации словаря межъязыковых синонимов на основе тезауруса Universal WordNet.

3. Разработан метод генерации искусственных атак «черного ящика» на языковые модели бинарной классификации, основанный на минимальных изменениях входных текстов на уровне букв и слов.

Практическая значимость. В ходе диссертационного исследования автором были разработаны программные средства обнаружения межъязыковых заимствований, исследованные на материале русского, английского, французского, испанского, немецкого и армянского языков. Указанные программные средства могут быть использованы в системах выявления плагиата в вузах и научных учреждениях. Разработанные программные инструменты были внедрены в Российско-Армянском Университете и в Высшем Аттестационном Комитете Республики Армения.

Достоверность и обоснованность научных результатов. В диссертационной работе корректно применяется математический аппарат, лежащий в основе используемых нейросетевых моделей, методов машинного обучения и средств автоматической обработки текстов; экспериментальные исследования разработанных методов проводятся на различных текстовых коллекциях с применением общепринятой методологии; полученные результаты экспериментов сопоставлены с известными лучшими

результатами и во многих случаях показано превосходство предложенных в диссертации решений.

Основные положения диссертации обсуждались на международных научных конференциях. Содержание исследования достаточно полно изложено в четырех печатных изданиях, в том числе в трех изданиях, индексируемых базой данных Scopus. Имеется свидетельство о государственной регистрации программы для ЭВМ.

Автореферат полностью соответствует содержанию диссертации и адекватно ее характеризует.

Замечания по диссертации.

1. В работе не указано, почему для решения задачи поиска кандидатов не были использованы многоязычные эмбединги (например, модели LabSE или E5).

2. Не расшифровано, что понимается под термином «малое количество данных» в определении малоресурсных языков на странице 6.

3. Непонятно, откуда взято ограничение на количество параметров модели (не более чем 300 миллионов), приведенное на странице 54.

4. Некоторые приведенные в диссертации результаты недостаточно проанализированы. Например, можно было провести анализ ошибок для результатов в таблице 33 на странице 107. Также желательно было исследовать статистическую значимость результатов в таблицах 25 и 30.

5. В работе присутствуют опечатки и пунктуационные ошибки.

Указанные замечания не влияют на общую положительную оценку диссертации.

Заключение. В целом, считаю, что диссертационная работа Аветисяна Карена Ишхановича соответствует требованиям пунктов 9–14 Положения о присуждении ученых степеней, утвержденного постановлением Правительства РФ от 24 сентября 2013 года № 842, предъявляемым к кандидатским диссертациям, а её автор заслуживает присуждения учёной степени кандидата технических наук по специальности 2.3.5 Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Официальный оппонент
профессор кафедры прикладной
математики и информатики
федерального государственного
бюджетного образовательного
учреждения высшего образования
«Вятский государственный
университет» (ФГБОУ ВО
«Вятский государственный
университет»),
доктор технических наук, доцент

14.11.2023г.

Котельников Евгений Вячеславович