

Отзыв научного руководителя

на диссертационную работу Аветисяна Карена Ишхановича на тему:

“Метод обнаружения межъязыковых заимствований в текстах”,

представленную на соискание ученой степени кандидата технических наук по специальности 2.3.5 - “Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей”.

Аветисян К.И. (1998 г.р.) окончил бакалавриат факультета “Прикладная Математика и Информатика” Российско-Армянского университета в Ереване в 2019 году. С 2019 года по 2021 год Аветисян К.И. учился в магистратуре Института математики и информатики Российско-Армянского университета по направлению “Машинное обучение”, с 2021 года там же проходит обучение в аспирантуре на кафедре системного программирования.

Студентом 3-го курса бакалавриата Аветисян К.И. поступил в лабораторию ИСП РАН в Российско-Армянском университете и с тех пор занимается вопросами разработки методов и программных инструментов для автоматической обработки текстов. Курсовая работа Аветисяна К.И. была посвящена построению моделей векторных представлений слов армянского языка. В бакалаврской дипломной работе исследовались методы разбиения слов для извлечения признаков в задаче морфологического анализа армянского языка. Магистерская диссертация была посвящена начальному этапу разработки методов обнаружения межъязыковых заимствований. В представляемой диссертации данная тема была развита и представляется разработанный итоговый метод обнаружения межъязыковых заимствований в текстах.

Тема диссертационного исследования Аветисяна К.И. имеет высокую актуальность в связи с отсутствием методов обнаружения межъязыковых заимствований применимых к большому количеству языков, в том числе малоресурсных. При этом университеты и другие организации Армении демонстрируют заинтересованность в проверке уникальности текстовых документов в контексте возможного наличия в них заимствований из документов других языков. В результате работы Аветисяна К.И. был разработан метод обнаружения межъязыковых заимствований превосходящий по качеству все существующие методы, а также применимый к большому количеству языков, в том числе малоресурсных.

Современные методы для решения исследуемой задачи опираются на использование инструментов машинного перевода, межъязыковых векторных представлений слов, мультязычных тезаурусах. Недостатком всех существующих методов является их слабая применимость к большому количеству языков, а в частности к малоресурсным языкам. В рамках работы Аветисяна К.И. предлагается метод основанный на

использовании межъязыкового словаря “синонимов” и модели бинарной классификации перевода между двумя предложениями, который применим к большому числу языков, в том числе к малоресурсным языкам таким как например, армянский язык.

При разработке метода обнаружения межъязыковых заимствований в текстах, также необходимо учитывать возможность атак на модели машинного обучения, с целью обмана системы. В связи с чем в рамках диссертационной работы был предложен новый метод “черного ящика” генерации атак на языковые модели бинарной классификации, превосходящий все существующие аналоги. Этот метод используется в предложенной Аветисяном К.И. методике выбора модели детального анализа с учетом риска осуществления атаки.

Результаты диссертации имеют существенную практическую значимость с точки зрения борьбы с межъязыковым плагиатом. Разработанная программная система, внедренная в Российско-Армянском университете, а также в Высший Аттестационный Комитет РА, позволит более строго отслеживать за качеством научных работ с точки зрения их уникальности. Кроме этого, представленные результаты могут быть полезны в различных задачах межъязыкового поиска для большого количества языков.

Считаю, что диссертация Аветисяна К.И. носит законченный характер и вносит значимый вклад в область обнаружения межъязыковых заимствований в документах большого количества языков, в том числе документах языков с ограниченными ресурсами. Его работа отвечает требованиям “Положения о порядке присуждения ученых степеней”, предъявляемых к кандидатским диссертациям по специальности 2.3.5 - “математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей”, а ее автор, Аветисян Карен Ишханович, заслуживает присуждения ему ученой степени кандидата технических наук.

Кандидат физико-математических наук,
заведующий отделом
Информационных систем ИСП РАН

Д.Ю. Турдаков

4 сентября 2023 г.