

**ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.1.120.01,
созданного на базе
Федерального государственного бюджетного учреждения науки
Институт системного программирования им. В.П. Иванникова
Российской академии наук
Министерства науки и высшего образования РФ
по диссертации на соискание ученой степени кандидата наук**

аттестационное дело № _____

решение диссертационного совета от 07 декабря 2023 года № 2023/14

О присуждении Аветисяну Карену Ишхановичу, гражданину Республики Армения (РА), ученой степени кандидата технических наук.

Диссертация «Метод обнаружения межъязыковых заимствований в текстах» по специальности 2.3.5 – «математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» принята к защите 06 октября 2023 г., протокол № 2023/09 диссертационным советом 24.1.120.01, созданным на базе Федерального государственного бюджетного учреждения науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ведомственная принадлежность: Министерство науки и высшего образования РФ; адрес: 109004, г. Москва, ул. А. Солженицына, дом 25), создан Приказом Минобрнауки России о советах по защите докторских и кандидатских диссертаций от 2 ноября 2012 г. № 714/нк.

Соискатель Аветисян Карен Ишханович, 1998 года рождения.

В 2021 году соискатель окончил Межгосударственное образовательное учреждение высшего образования «Российско-Армянский Университет». В настоящее время соискатель обучается в аспирантуре Межгосударственного образовательного учреждения высшего образования «Российско-Армянский Университет».

К.И. Аветисян работает научным сотрудником в Центре передовых программных технологий Межгосударственного образовательного учреждения

высшего образования «Российско-Армянский Университет» (ведомственная принадлежность: Министерство науки и высшего образования РФ и Министерство образования и науки РА).

Диссертация выполнена на кафедре системного программирования Института математики и информатики Межгосударственного образовательного учреждения высшего образования «Российско-Армянский Университет» (ведомственная принадлежность: Министерство науки и высшего образования РФ и Министерство образования и науки РА).

Научный руководитель – кандидат физико-математических наук Турдаков Денис Юрьевич, ФГБУН Институт системного программирования им. В.П. Иванникова Российской академии наук, Отдел информационных систем, заведующий отделом.

Официальные оппоненты:

1. Котельников Евгений Вячеславович, доктор технических наук, доцент, профессор кафедры прикладной математики и информатики Федерального государственного бюджетного образовательного учреждения высшего образования «Вятский государственный университет»,
2. Чехович Юрий Викторович, кандидат физико-математических наук, исполнительный директор АО «Антиплагиат»

дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет им. М. В. Ломоносова» (г. Москва) в своем положительном заключении, подписанном Тихонравовым Александром Владимировичем (заведующий лабораторией вычислительного эксперимента и моделирования Научно-исследовательского вычислительного центра Московского государственного университета имени М.В. Ломоносова, доктор

физико-математических наук, профессор) и Воеводиным Владимиром Валентиновичем (директор Научно-исследовательского вычислительного центра Московского государственного университета имени М.В. Ломоносова, член-корреспондент РАН, доктор физико-математических наук, профессор), указала, что диссертационная работа содержит новые научные результаты, имеющие существенное значение для науки и практики.

Соискатель имеет 9 опубликованных работ, в том числе по теме диссертации опубликовано 4 работ, из них в рецензируемых научных изданиях опубликовано 3 работ. Ниже приведена краткая характеристика наиболее значимых работ соискателя:

1) Avetisyan K., Gritsay G., Grabovoy A. Cross-Lingual Plagiarism Detection: Two Are Better Than One //Programming and Computer Software. – 2023. – Т. 49. – №. 4. – С. 346-354. - Scopus. В статье представляется метод улучшения результатов обнаружения межъязыковых заимствований путем последовательного объединения двух различных методов. Личный вклад соискателя заключается в реализации объединения, планировании экспериментов, участии в их проведении и обработке результатов, а также написании текста публикации.

2) Ter-Hovhannisyanyan T., Avetisyan K. Transformer-Based Multilingual Language Models in Cross-Lingual Plagiarism Detection //2022 Ivannikov Memorial Workshop (IVMEM). – IEEE, 2022. – С. 72-80. - Scopus. В статье проводятся эксперименты по выявлению лучшей трансформер-основанной модели для задачи определения перевода между двумя предложениями. Личный вклад соискателя заключается в планировании экспериментов, участии в их проведении и обработке результатов, а также написании текста публикации.

3) Ghukasyan T., Yeshilbashyan Y., Avetisyan K. Subwords-only alternatives to fastText for morphologically rich languages //Programming and Computer Software. – 2021. – Т. 47. – С. 56-66. - Scopus. В статье представлено решение проблемы разреженности данных в методах лемматизации и морфологического анализа языков с богатой морфологией (армянский в том числе). Личный вклад

соискателя заключается в реализации новых методов векторного представления слов, участии в проведении экспериментов и написании текста публикации.

4) Malajyan A., Avetisyan K., Ghukasyan T. ARPA: Armenian Paraphrase Detection Corpus and Models // 2020 Ivannikov Memorial Workshop (IVMEM). — 2020. — С. 35—39. Для армянского языка впервые предложен программный инструмент, решающий эту задачу, а также полуавтоматический универсальный метод генерации парафразов предложений, существенно сокращающий необходимость привлечения экспертов в процессе разметки. Соискатель принимал непосредственное участие в разработке программного инструмента и метода генерации, в сборе, подготовке и ручной разметке примеров, планировании и проведении экспериментов.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- разработан новый метод обнаружения межъязыковых заимствований применимый к задаче обнаружения межъязыковых заимствований в текстах малоресурсных языков;
- разработан новый метод генерации словаря “межъязыковых” синонимов, основанный на использовании тезауруса Universal WordNet и машинного перевода;
- разработан алгоритм генерации состязательных атак на языковые модели бинарной классификации, превосходящий по доле успешных атак существующие аналоги;
- предложена методика выбора языковой модели для одного из этапов обнаружения межъязыковых заимствований, с учетом рисков быть подвергнутым состязательным атакам.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

- разработан и внедрен в учебных учреждениях Республики Армения новый метод обнаружения межъязыковых заимствований применимый к малоресурсным языкам;

- разработан новый метод генерации словаря межъязыковых синонимов, позволяющего достичь высоких показателей метрики полноты для этапа извлечения кандидатов в задаче обнаружения межъязыковых текстовых заимствований;
- разработан новый метод генерации состязательных атак “черного ящика” на языковые модели бинарной классификации, превосходящий по доле успешных атак, а также по дистанции Левенштейна и семантической близости существующие аналоги.
- разработана методика выбора языковой модели для этапа детального анализа, учитывающая угрозу возможности осуществления состязательных атак;
- Созданы новые тестовые наборы данных для 5 языков для задачи обнаружения межъязыковых текстовых заимствований.

Оценка достоверности результатов исследования выявила:

- принятые в диссертации решения базируются на анализе практики;
- полученные результаты подтверждаются экспериментами;
- эксперименты были проведены на существующих тестовых наборах данных;
- метод внедрен в образовательных учреждениях Республики Армения.

Личный вклад соискателя: все представленные на защиту решения получены соискателем лично.

В ходе защиты диссертации были высказаны следующие критические замечания:

- не объяснено, откуда взято ограничение на количество параметров модели бинарной классификации (не более чем 300 миллионов);
- в работе не указано, почему для решения задачи поиска кандидатов не были использованы многоязычные эмбединги.

Соискатель Аветисян Карен Ишханович согласился с замечаниями, ответил на задаваемые ему в ходе заседания вопросы.

На заседании 07 декабря 2023 г. диссертационный совет принял решение присудить Аветисяну К.И. ученую степень кандидата технических наук за решение научной задачи, имеющей значение для развития соответствующей отрасли знаний.

При проведении тайного голосования диссертационный совет в количестве 15 человек, из них 7 докторов наук по специальности рассматриваемой диссертации, участвовавших в заседании, из 22 человек, входящих в состав совета, проголосовали: за – 15, против – 0.

Заместитель председателя диссертационного совета,
доктор физико-математических наук

Петренко А. К.

Ученый секретарь диссертационного совета,
кандидат физико-математических наук

Зеленов С. В.

07 декабря 2023 г.