

## **ОТЗЫВ**

официального оппонента на диссертацию

Аветисяна Карена Ишхановича

«Метод обнаружения межъязыковых заимствований в текстах»,

представленную к защите на соискание ученой степени

кандидата технических наук по специальности

2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

### **Актуальность темы**

Диссертация Аветисяна К. И. посвящена методу обнаружения межъязыковых текстовых заимствований. Раскрытие заимствованных работ выполняет важную роль в образовательных и научных организациях, так как помогает контролировать качество работ учащихся, статей и других публикаций. В рамках данной диссертационной работы рассматриваются заимствования, производимые из текстов, язык которых отличен от языка текста, в котором ищутся заимствования. В работе представляется метод обнаружения межъязыковых заимствований применимый к большому количеству языков (в том числе, малоресурсных). Низкое качество методов применимых к большому количеству языков актуализирует представляемую работу.

### **Практическая значимость работы**

Основная практическая значимость диссертации заключается в разработке методов и программных средств обнаружения межъязыковых заимствований, применимых в учебных заведениях различных стран, в том числе тех, официальным языком которых является малоресурсный язык. Программные средства были опробованы, например, на таком малоресурсном языке как армянский, и показали высокие результаты. Исходный метод был расширен с внедрением дополнительного этапа детального анализа и специального алгоритма, предназначенного для объединения обнаруженных текстовых фрагментов. Эти усовершенствования были внедрены в программную систему и привели к увеличению эффективности обнаружения случаев межъязыковых

заимствований. В результате этих изменений, усовершенствованный метод теперь может быть рассмотрен как новый подход для выявления межъязыковых заимствований с более высокой эффективностью.

Были созданы тестовые наборы данных для обнаружения случаев межъязыкового плагиата в двух разных сценариях: в первом сценарии в пределах одного анализируемого документа может присутствовать заимствование из текстов на разных языках, а во втором сценарии в пределах одного документа могут быть обнаружены заимствования только из текстов на одном языке

Также, представлена методика отбора языковой модели, которая будет использоваться в процессе детального анализа, с учетом риска возможных атак, направленных на нарушение системы.

### **Научная новизна работы**

Научная новизна работы заключается в разработке метода обнаружения межъязыковых заимствований не использующего инструменты разрешения лексической неоднозначности слов и машинного перевода. Неиспользование подобных инструментов позволяет представленному методу быть применимым к большому количеству языков, в том числе малоресурсных. Дополнительно, представляемый метод обходит по точности все инструменты, представленные в открытом доступе.

В рамках работы был разработан метод генерации состязательных атак на языковые модели, превосходящий по доле успешных атак, а также по семантической близости сгенерированных примеров с их оригиналами, все существующие методы генерации состязательных атак на языковые модели бинарной классификации.

### **Обоснованность и достоверность результатов диссертации**

Результаты диссертации являются новыми, их достоверность не вызывает сомнений. Ошибок в выводах и постановках экспериментов не обнаружено. Все полученные результаты подтверждаются экспериментами.

### **Структура и содержание диссертации**

Диссертация состоит из введения, четырёх глав, заключения и двух приложений. Библиография включает 111 наименований.

Во введении обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, формулируется цель, ставятся задачи работы, сформулированы научная новизна и практическая значимость представляемой работы.

Первая глава посвящена исследованию литературы по изучаемой проблеме. Производится обзор, а также выделяются ключевые аспекты различных работ по рассматриваемой тематике.

Вторая глава посвящена описанию метода обнаружения межъязыковых заимствований, не использующего специализированных для конкретных языков инструментов, и примененного для большого количества языков. В главе рассматриваются два основных этапа: этап извлечение кандидатов, основанный на использовании словаря межъязыковых синонимов и этап детального анализа, основанный на бинарной классификации перевода. Дополнительно, в главе представляется новый метод генерации состязательных атак на языковые модели бинарной классификации. Представляется методика выбора модели классификации перевода с учетом вероятности быть подвергнутым состязательным атакам.

В третьей главе рассматриваются различные существующие, а также новые, сгенерированные тестовые выборки, на которых и была произведена оценка работы метода. В главе представляются итоговые результаты, а также производится сравнительный анализ с другими существующими в открытом доступе системами обнаружения межъязыковых заимствований.

Четвертая глава посвящена способам слияния представляемого в диссертации метода с методом, представленным в статье Vakhteev et al., 2019. Представляются результаты, полученные слитным методом, сильно улучшающие метрики обнаружения межъязыковых заимствований.

#### **Рекомендации по использованию результатов диссертации**

Разработанный алгоритм генерации словаря межъязыковых синонимов, а также его применение в межъязыковом поиске, может быть применено в

различных системах множества других языков. Тестовые наборы, сгенерированные в рамках работы, могут быть использованы для сравнения различных систем обнаружения межъязыковых заимствований. Кроме того, представленный подход допускает масштабирование, что позволяет рассчитывать в перспективе на повышение качества решаемых задач путем перенастройки алгоритма на расширенных наборах данных. Дополнительно, представленный метод генерации состязательных атак может быть использован для проверки различных систем на уязвимость.

### **Апробация и публикация результатов диссертации**

Результаты диссертационной работы докладывались на международных и локальных научных конференциях. Опубликовано три научные работы в журналах, индексируемых в Scopus, а также одно свидетельство о государственной регистрации программы для ЭВМ.

### **Недостатки в содержании и оформлении диссертации**

По диссертации имеются следующие замечания:

1. Помимо оценок качества точности представленного метода, также было бы полезно представить оценки его производительности и скорости обнаружения заимствований.

2. Было бы желательно для малоресурсных языков сравнить представляемый метод с методом, использующим машинный перевод для обнаружения межъязыковых заимствований.

Указанные замечания не умаляют достоинств работы. Диссертация является законченной научно-квалификационной работой и оформлена в соответствии с требованиями, установленными Минобрнауки России. Автореферат диссертации адекватно отражает содержание основной работы.

### **Заключение**

Таким образом, диссертация Аветисяна К. И. «Метод обнаружения межъязыковых заимствований в текстах» является самостоятельным, завершенным научным исследованием и полностью соответствует требованиям п. 9 «Положения о порядке присуждения ученых степеней», утвержденного

постановлением Правительства Российской Федерации от 24.09.2013 г. № 842, а Аветисян Карен Ишханович заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент,

Чехович Юрий Викторович,  
кандидат физико-математических наук,  
исполнительный директор АО «Антиплагиат».

Дата «20» ноября 2023 г.