

**УДК 004.89**

**ИССЛЕДОВАНИЕ МЕТОДОВ АВТОМАТИЗАЦИИ ФОРМАЛИЗАЦИИ ТРЕБОВАНИЙ К ПРОГРАММНОМУ ОБЕСПЕЧЕНИЮ**

**Лисенков М.А.** федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

**Научный руководитель – к.т.н., доцент Осипов Н.А.**  
федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

В статье проводится анализ состояния исследований в сфере инженерии требований, и изучение возможности решения задач с помощью инструментария обработки естественного языка — подраздела машинного обучения, работающего с текстовыми данными (NLP4RE). Отмечаются существующие проблемы, такие как недостаток валидированных и оцениваемых исследований, недостаток решений задач поздних стадий разработки требований, малое количество данных для обучения моделей обработки естественного языка. Кроме того, проводится анализ существующих подходов обработки естественного языка, используемых для решения задач инженерии требований, где в особенности выделяются различные предобученные модели, которые не затрагиваются недостатком данных для обучения в предметной области.

**Введение.**

В области инженерии требований, являющейся неотъемлемой частью процесса проектирования ПО, существует ряд проблем связанных с природой естественного языка, а также проблемами коммуникации между разработчиками и заинтересованными сторонами. С развитием технологий обработки естественного языка появились исследования решения проблем инженерии требований с помощью этих технологий.

Целью данного исследования является изучение текущего положения в сфере обработки естественного языка для инженерии требований. В рамках исследования рассматриваются следующие вопросы:

- Решение каких проблем наиболее актуально в рамках предметной области инженерии требований?
- Какие подходы обработки естественного языка более эффективны для решения проблем инженерии требований?

**Основная часть.**

На данный момент в области NLP4RE с помощью инструментов обработки естественного языка решаются задачи всех этапов инженерии требований, включающие в себя: создание различных моделей на основе документов инженерии требований, обнаружение недостатков в документах инженерии требований, выделение ключевых сущностей, классификация требований, определение и отслеживание изменений связей между артефактами инженерии требований, а также поиск артефактов инженерии требований в существующих репозиториях.

При этом большая часть исследований посвящены решению проблем ранних стадий инженерии требований: моделированию, обнаружению недостатков и выделению ключевых сущностей. Кроме того, существующие исследования направлены на изучение отдельных проблем без ориентации на модульность и создание

единой системы инженерии требований на основе технологий разработки естественного языка.

Проблема входных и обрабатываемых данных существующих исследований, заключается в том, что большая часть исследований использует в этом качестве уже формализованные документы инженерии требований, при том, что более актуальна работа именно с документами, написанными на естественном языке, например с результатами опросов и интервью заинтересованных сторон или пользовательских историй. Также существует проблема малого количества текстовых данных, которые могут быть недостаточны для обучения некоторых моделей обработки естественного языка.

Кроме того, большая часть исследований в сфере на данный момент являются case-study, при этом количество оценочных и валидированных метриками исследований составляет относительно небольшую часть.

Для решения проблем инженерии требований широко могут использоваться различные подходы разработки естественного языка подходы: POS-tagging, Семантический парсинг, NER, Векторные модели, а также различные модели машинного обучения.

При этом некоторые модели сталкиваются с недостатком наборов данных для обучения в предметной сфере. Эта проблема может быть решена использованием предобученных моделей машинного обучения, такие как BERT, показавший большую эффективность в сравнении с векторными моделями в существующих исследованиях, или UniLM.

**Выводы.** Данное исследование выделяет ключевые проблемы как сферы NLP4RE, так и часто используемых подходов обработки естественного языка. В качестве продолжения исследований предлагается проведение сравнительного исследования различных моделей обработки естественного языка для решения проблем ранних стадий инженерии требований в работе с документами инженерии требований, основанных на естественном.

Лисенков М.А. (автор)

Подпись

Осипов Н.А. (научный руководитель)

Подпись