

**УТВЕРЖДАЮ**

Ректор автономной некоммерческой  
организации высшего образования  
«Университет Иннополис»  
д. физ.-мат. наук, проф. А. Г. Тормасов

---

« 18 » \_\_\_\_\_ мая \_\_\_\_\_ 2021 г.

### **ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ**

**на диссертацию Алимовой Ильсеяр Салимовны «Нейросетевой механизм кросс-внимания в задачах извлечения информации из текстов на примере биомедицинских данных», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».**

Диссертация Алимовой И.С. посвящена методам и моделям на основе нейронных сетей с механизмом кросс-внимания для извлечения информации из текстов. Извлечение информации из неструктурированных текстов является одной из важнейших задач современного интеллектуального анализа текстов и имеет практическое применение для широкого круга прикладных задач и в особенности в сфере биомедицины. Таким образом, подтверждается актуальность диссертационной работы. Работа соответствует заявленной специальности 05.13.11.

Структура диссертации построена следующим образом. В первой главе дается обзор литературы по теме диссертации: методы работы с текстом, классификация текста, извлечение отношений, извлечение информации из текстов биомедицинской тематики, а также перечислены недостатки существующих методов, разработанных для решения поставленных задач.

Во второй главе приводится формальная постановка задачи классификации текста, описание предложенного метода на основе нейронной сети с долгой краткосрочной памятью и механизмом кросс-внимания с набором признаков, оценка информативности признаков, описание наборов данных, на которых оценивались разработанные методы, постановка экспериментов, результаты, выводы. Результаты экспериментов показывают высокое качество разработанной модели в задаче классификации на пяти англоязычных и одном русскоязычном корпусах текстов биомедицинской тематики. Дополнительно автор приводит качественный анализ результатов на примере задачи извлечения побочных эффектов о лекарственных препаратах из текстов отзывов пользователей на русском языке.

В третьей главе рассматривается задача извлечения отношений между двумя сущностями. Как и во второй главе даны формальная постановка задачи, описание разработанной модели, корпусов, на которых проводилась оценка предложенной модели, результаты и выводы. В качестве метода для извлечения отношений была также предложена нейронная сеть на основе долгой краткосрочной памяти и механизма кросс-

внимания с дополнительным входом для второй сущности. Результаты экспериментов показали, что предложенная модель показывает более высокое качество по сравнению с существующими современными языковыми моделями при кросс-доменной проверке. Кроме того, автором была предложена модель с набором признаков и проведена оценка информативности предложенных признаков. Также автором был проведен анализ векторных представлений контекста между сущностями, который показывает, что предложенная модель способна лучше моделировать контексты для корпусов из разных источников.

В четвертой главе приводится описание разработанного программного обеспечения для решения задач классификации сущностей и извлечения отношений. Автором описаны детали архитектуры разработанной системы, модули, а также примеры консольных команд для работы с программой.

Таким образом, основными результатами диссертационной работы являются:

- новый метод классификации сущностей на основе модели нейронной сети с долгой краткосрочной памятью и механизмом кросс-внимания с набором признаков, извлеченных из текста;
- новый метод извлечения отношений между сущностями, основанный на модели нейронной сети с долгой краткосрочной памятью и механизмом кросс-внимания, в которой имеются отдельные входы для сущностей и контекста между сущностями;
- программные реализации разработанных методов.

Указанные результаты являются новыми, их достоверность подтверждена. В работе имеются как прикладные, так и теоретические результаты. Практическая применимость результатов не вызывает сомнений и доказана в самой работе. Ошибок в постановках экспериментов и выводах не обнаружено.

В качестве сильной стороны диссертации хочется отметить, что разработанные методы являются универсальными для русского и английского языков, протестированы на большом количестве корпусов, состоящих из текстов, собранных с различных ресурсов, а также имеют более высокую возможность практического применения за счет высоких результатов при кросс-доменном тестировании.

В работе можно отметить следующие недостатки. В задаче извлечения отношений приводится описание генерации отрицательных примеров, однако отсутствует статистика по количеству сгенерированных примеров для каждого корпуса. В задаче классификации сущностей все англоязычные модели принимают на вход вектора, полученные из модели векторного представления слов, предобученной на текстах социальных медиа. Хотелось бы видеть, на сколько изменятся результаты, если на вход модели подавать вектора, предобученные на текстах более близкого к текстам корпуса домена. Общие принципы механизма внимания необходимо было пояснить в общем обзоре, в котором рассматривались различные архитектуры нейронных сетей. В таблице 25 проведено сравнение предложенных в работе методов между собой без упоминания других моделей из таблицы 20, которые имеют более качественные результаты работы, из-за чего оценка в пункте 3.6.4 является не показательной.

Имеется ряд орфографических опечаток, например: на странице 15 во фразе «Векторное представление обучатся» опечатка в слове обучается, на странице 18 во фразе «Набор правил по которым» пропущена запятая, на странице 113 в уточнении «(см. секции ??)» вместо ссылки на секцию знаки вопроса и т.д. Однако, в основном, эти недочеты незначительны.

Тем не менее, указанные замечания не ставят под сомнение ценность основных результатов работы. Диссертация является законченной научно-квалификационной работой, выполненной автором самостоятельно на высоком уровне. Основные этапы работы, результаты и выводы отражены в автореферате. В частности, прикладные результаты работы могут быть использованы в программных системах и продуктах для интеллектуального анализа текстов биомедицинской тематики и извлечения дополнительной информации о побочных эффектах лекарственных препаратов.

Таким образом, диссертационная работа Алимовой Ильсеяр Салимовны «Нейросетевой механизм кросс-внимания в задачах извлечения информации из текстов на примере биомедицинских данных» удовлетворяет всем требованиям ВАК, предъявляемым к кандидатским диссертациям, а автор работы заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.11 - математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Диссертация и отзыв обсуждены на заседании ученого совета Автономной некоммерческой организации высшего образования «Университет Иннополис» (Протокол № УС-1/2021/05/18 от 18 мая 2021 года).

Сведения о ведущей организации: Автономная некоммерческая организация высшего образования «Университет Иннополис».

Адрес: 420500, г. Иннополис, ул. Университетская, д.1

Электронная почта: [university@innopolis.ru](mailto:university@innopolis.ru); [university.innopolis.ru](http://university.innopolis.ru)

Телефон: +7 (843) 203-92-53

ОКПО 26762138; ОГРН 1121600006142;

ИНН/КПП 1655258235/161501001

Ректор

Доктор физико-математических наук,  
профессор

Тормасов Александр Геннадьевич