

## ОТЗЫВ

научного руководителя

о кандидатской диссертации Тихомирова М.М.

*"Методы автоматизированного пополнения графов знаний"  
на основе векторных представлений"*,

представленной на соискание ученой степени кандидата

физико-математических наук

по специальности 05.13.11 «Математическое и программное обеспечение  
вычислительных машин, комплексов и компьютерных сетей»

Одним из значимых направлений в современных исследованиях в сфере автоматической обработки текстов является поиск комбинированных методов, позволяющих сочетать мощность нейронных сетей при обучении на конкретных наборах данных и обобщающие возможности знаний о языке и мире. Такие знания обычно представляются в специализированных ресурсах, например, графах знаний – семантических сетях, содержащих как классификацию понятий и конкретных сущностей (таксономию), так и большие объемы описаний конкретных (именованных) сущностей и отношений между ними. Создание и развитие графов знаний требует постоянного их пополнения, источником которого могут являться текстовые коллекции. Однако в настоящее время качество извлечения знаний, в частности, таксономических отношений, из текстов остается не достаточно высоким, на текущий момент невозможно автоматически создавать семантические ресурсы ручного качества.

В работе Тихомирова М.М. предложен подход к извлечению таксономических отношений из текстовых коллекций на основе мета-векторных представлений, построенных посредством применения нейронных сетей – автокодировщиков. В работе показано, что метод позволяет комбинировать векторные представления слов различной природы, а также векторные представления графов, что дает возможность улучшения качества предсказания таксономических отношений для слов по сравнению с

базовыми векторными представлениями.. Результаты подхода протестированы на разных наборах данных, для русского и английского языков

Поскольку графы знаний содержат и описания большого числа именованных сущностей, то вторым направлением исследований стали методы извлечения именованных сущностей в конкретных предметных областях. Был предложен подход, сочетающий два вида размеченных наборов данных (общий и специализированный), автоматически размеченные данные, полученные на основе нового метода псевдоразметки, предложенного в работе, а также дообучения нейроязыковой модели BERT на текстовой коллекции предметной области.

Предложен специализированный интерфейс для автоматизированного пополнения графа знаний, в котором совмещены как наиболее вероятные предсказанные гиперонимы для новых понятий, так и возможность просмотра сети графа знаний для корректного ввода пополнения.

При выполнении диссертационной работы М.М. Тихомиров проявил себя зрелым исследователем, способным как к самостоятельному выдвижению новых идей при решении задач в сложной области, так и к их реализации и проверке.

Полученные в диссертации результаты являются новыми, их достоверность подтверждается экспериментами на разных наборах данных (для русского и английского языков, данных общей направленности и конкретной предметной области), а также публикацией созданных моделей и наборов данных.

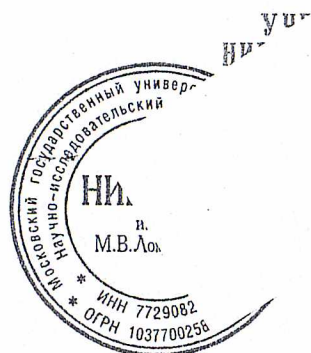
Диссертационная работа выполнена на высоком уровне, представляет собой законченное научное исследование и удовлетворяет всем требованиям, предъявляемым к кандидатским диссертациям. Ее результаты получены лично автором и прошли квалифицированную апробацию. Автореферат адекватно отражает содержание диссертации. Рекомендую присудить ее автору М.М. Тихомирову ученую степень кандидата физико-математических

наук по специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Лукашевич Наталья Валентиновна,  
Доктор технических наук,  
Ведущий научный сотрудник,  
НИВЦ МГУ имени М.В. Ломоносова  
(119991, Москва, Ленинские горы, д.1, стр. 4)  
Телефон: +7(926)1446163  
Адрес электронной почты: [louk\\_nat@mail.ru](mailto:louk_nat@mail.ru)

20.04.2022

*Подпись Н.В. Лукашевич заверяю*



УЧ  
НИ  
РЕТАРЬ  
ЛОМОНОСОВА  
ВОРДВ