

**ОТЗЫВ**  
**о выпускной квалификационной работе Мамаева Ивана Дмитриевича**  
**«АВТОМАТИЧЕСКОЕ ВЫЯВЛЕНИЕ СКРЫТЫХ СООБЩЕСТВ В ТЕКСТАХ РУССКОЯЗЫЧНОГО**  
**КОРПУСА СОЦИАЛЬНЫХ СЕТЕЙ»**

Исследование Ивана Дмитриевича Мамаева проведено на стыке корпусной лингвистики, социолингвистики, анализа данных. Внимание автора было сконцентрировано на социальных сетях как объекте лингвистического анализа и компьютерного моделирования. В фокусе ВКР оказались скрытые сообщества, неявные группы пользователей, которые формируются не по наличию непосредственных связей, а по общности интересов, членству в пользовательских группах, кросспостингу и т.д. С одной стороны, специалисты по социальным медиа уже разработали математические методы выявления таких сообществ, однако они не учитывают лингвистическую природу наполнения социальных сетей. С другой стороны, есть объективные лингвистические параметры, по которым можно определить принадлежность пользователей к одному сообществу. Иван Дмитриевич в своем проекте предложил авторское решение проблемы выявления скрытых сообществ, являющееся альтернативой математических подходов. Это решение, совмещающее математическое моделирование и лингвистическую экспертизу, основано на применении мультимодальной тематической модели, которая учитывает авторство сообщений и их тематику, а также допускает учет лексических конструкций в корпусе и обобщение тем – автоматическое назначение меток тем.

Эксперименты, проведенные Иваном Дмитриевичем, имеют ряд особенностей. Прежде всего, автор ВКР самостоятельно собрал и обработал репрезентативный корпус постов из социальной сети ВКонтакте, используя не стандартное программное обеспечение, а инструменты, разработанные им самим. Выбранная для выявления скрытых сообществ автор-тематическая модель строилась на основе алгоритмов библиотеки *genism*, которые были существенно доработаны Иваном Дмитриевичем. Автор предложил новаторские решения на этапе сборки тем с учетом конструкций и на этапе обобщения тем с помощью меток. Иван Дмитриевич реализовал два метода назначения меток тем с использованием внешних источников данных – графовый метод извлечения меток из выдачи поисковой машины с ранжированием по PageRank и фильтрацией по лексико-грамматическим шаблонам, словарный метод поиска родовых терминов в *wordnet*-тезаурусе с использованием модели векторных вложений слов *word2vec*. Все полученные данные были детально проанализированы и проинтерпретированы, и для оценки качества построенной модели была проведена экспертная оценка назначенных меток.

Проект Ивана Дмитриевича отличается масштабностью и новизной. Предложенная им разработка может рассматриваться как задел для исследования более высокого уровня. Результаты компьютерного моделирования и анализа данных отражены в ВКР с подробными комментариями, а также представлены в представительных приложениях. Все задачи решены успешно и в полном объеме. Результаты работы обсуждались на ведущих международных конференциях по искусственному интеллекту и представлены в ряде публикаций, индексируемых в наукометрической базе данных Scopus. Научный руководитель считает должным отметить, что все исследовательские процедуры были выполнены Иваном Дмитриевичем самостоятельно и на высочайшем профессиональном уровне, результаты изложены четко, ясно и убедительно. В процессе работы над ВКР автор проявляет трудолюбие, настойчивость, пытливость, открытость новой информации. Хочется отметить, что за время обучения в магистратуре и в ходе подготовки ВКР Иван Дмитриевич постоянно совершенствовался, он является опытным специалистом в области автоматической обработки данных, программирования, статистической обработки результатов наблюдений и т.д. Иван Дмитриевич – молодой исследователь с широким научным кругозором. Совместная работа с Иваном Дмитриевичем приносит радость.

Итак, выпускная квалификационная работа, представленная к защите, соответствует всем требованиям, предъявляемым к научным произведениям такого уровня, а ее автор заслуживает присвоения квалификации магистра лингвистики. Содержание ВКР полностью соответствует заявленной в названии теме, тема раскрыта полностью, структура ВКР соответствует цели и задачам исследования, в ВКР в полной мере освещены актуальные проблемы теоретического и практического характера. При выполнении исследования и подготовке текста ВКР автор использовал современную литературу, достижения науки и практики. Текст ВКР прекрасно структурирован, написан строгим научным стилем, автор четко формулирует теоретические предпосылки исследования, детально комментирует практический материал, дает развернутое обоснование выводов. Результаты, достигнутые автором ВКР, имеют высокую теоретическую значимость, практическую ценность и могут найти применение в самых разных областях прикладной лингвистики.

11 июня 2021 г.



Научный руководитель, кандидат филологических наук, доцент кафедры  
математической лингвистики СПбГУ О.А.Митрофанова

### Согласие на обработку персональных данных

Я, **Митрофанова Ольга Александровна**, даю согласие на обработку своих персональных данных оператору - Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет» (далее - СПбГУ), 199034, Санкт-Петербург, Университетская наб., д. 7-9, на следующих условиях:

1. Оператор осуществляет обработку персональных данных исключительно в связи с осуществлением рецензирования и проведения защиты выпускных квалификационных работ обучающихся СПбГУ в целях реализации принципа открытости образовательной деятельности.
2. Перечень персональных данных, передаваемых Оператору на обработку:
  - фамилия, имя, отчество;
  - место работы, должность;
  - ученая степень и звание (при наличии);
  - контактный телефон и адрес электронной почты.
3. Оператор имеет право на обработку персональных данных, то есть совершение, в том числе, следующих действий: обработку (включая сбор, систематизацию, накопление, хранение, уточнение (обновление, изменение), использование, обезличивание, блокирование, уничтожение персональных данных).
4. Данным заявлением разрешаю считать общедоступными, в том числе выставлять в сети Интернет, следующие персональные данные: фамилия, имя, отчество, место работы, должность, ученая степень и звание (при наличии).
5. Обработка персональных данных осуществляется оператором в соответствии с нормами Федерального закона от 27.07.2006 № 152-ФЗ «О персональных данных» и смешанным способом.
6. Срок действия данного Согласия не ограничен.

11 июня 2021 г.

Митрофанова О.А.