

УТВЕРЖДАЮ:

Проректор по научной работе  
Санкт-Петербургского  
государственного университета



/ С. В. Микушев/

2024 г.

### ЗАКЛЮЧЕНИЕ

#### федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет»

Диссертация И. Д. Мамаева «Лингвистическое исследование скрытых сообществ в корпусе социальных медиа с применением мультимодальных тематических моделей» выполнена на кафедре математической лингвистики Федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет».

В 2021 г. окончил магистратуру по направлению 45.04.02 «Компьютерная и прикладная лингвистика» в Федеральном государственном бюджетном образовательном учреждении высшего образования «Санкт-Петербургский государственный университет». Присуждена степень «магистр».

В период с 2021 по 2024 гг. обучался в аспирантуре Федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» по направлению подготовки 45.06.01 Языкознание и литературоведение, присвоена квалификация «Исследователь. Преподаватель-исследователь».

Диплом об окончании аспирантуры АС № 03028 от 15.07.2024 года, выдан Федеральным государственным бюджетным образовательным учреждением высшего образования «Санкт-Петербургский государственный университет».

С 2021 г. работает преподавателем кафедры Р7 «Теоретической и прикладной лингвистики» федерального государственного бюджетного образовательного учреждения высшего образования «Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова».

С 2022 г. работает ассистентом на кафедре математической лингвистики федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» (по совместительству).

**Научный руководитель – Митрофанова Ольга Александровна**, кандидат филологических наук, доцент, доцент кафедры математической лингвистики федерального

государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет».

По итогам обсуждения принято следующее заключение.

### **Оценка выполненной соискателем работы**

В диссертации И. Д. Мамаева «Лингвистическое исследование скрытых сообществ в корпусе социальных медиа с применением мультимодальных тематических моделей» решена научная задача построения модели скрытых сообществ методами тематического моделирования, а также создания лингвистических профилей пользователей для каждого сообщества. В результате разработан новый подход к автоматическому выявлению и лингвистической интерпретации скрытых сообществ в социальных медиа. На основе данного подхода был разработан представительный русскоязычный корпус с разметкой тематики постов отдельных авторов и их сообществ, а также сформированы лингвистические профили скрытых сообществ, включающие лексические, морфологические и синтаксические характеристики текстов.

Диссертация И. Д. Мамаева «Лингвистическое исследование скрытых сообществ в корпусе социальных медиа с применением мультимодальных тематических моделей» представляет собой самостоятельное, целостное, завершенное квалификационное научное исследование, имеющее теоретическое и практическое значение. Работа отвечает требованиям, предъявляемым к диссертациям, представленным на соискание степени кандидата филологических наук по специальности 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика.

Диссертация может быть представлена в совет по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук.

### **Личное участие соискателя в получении результатов, изложенных в диссертации**

В процессе работы над диссертацией И. Д. Мамаев провел самостоятельное исследование, в ходе которого были получены следующие результаты: 1) обобщен исследовательский опыт по изучению интернет-дискурса, структур сетевых сообществ и инструментария компьютерной лингвистики; 2) собраны данные из открытого доступа сети Интернет для создания корпуса пользовательских постов; 3) разработаны процедуры предобработки постов и решения возникших проблем; 4) обоснован выбор алгоритма тематического моделирования как лингвистического метода создания модели скрытых сообществ; 5) детализированы структуры скрытых сообществ с привлечением социоматематических методов; 6) рассчитаны внутритекстовые корреляты, характеризующие скрытое сообщество пользователей.

**Основные научные результаты**, достигнуты лично И. Д. Мамаевым в ходе диссертационного исследования: добавление параметра авторства в тематическую модель позволило выявить темы узкой семантики; установлено, что отсутствие лингвистического анализатора в математических алгоритмах влияет на итоговое количество выделенных сообществ; показано, что посты пользователей, объединенных единой темой, наиболее полно характеризуются с точки зрения морфологии и синтаксиса; значимые лексические корреляции представлены в небольшом количестве профилей. Это позволило выделить в корпусе текстов 23 скрытых сообщества разной тематической направленности.

### **Степень достоверности результатов исследования**

Степень достоверности выводов обеспечивается хорошо проработанной теоретической базой исследования, формированием оригинальной концепции исследования и ее последовательной реализацией, профессионально проведенным лингвистическим анализом и результатами компьютерной реализации процедуры лингвистического профилирования. Достоверность практических результатов исследования обеспечивается репрезентативностью эмпирических данных, количественными оценками качества обучения тематических моделей, применением методов оценки согласованности экспертов при разметке тематических моделей, что позволило создать модель скрытых сетевых сообществ, а также использованием методов статистических исследований, что позволило представить только значимые корреляты на морфологическом, синтаксическом и лексическом уровнях.

### **Новизна результатов проведенного исследования**

Научная новизна диссертации заключается в следующем:

- разработана лингвистическая процедура создания и описания скрытых сообществ;
- впервые для построения модели скрытых сообществ применяется мультимодальный подход в тематическом моделировании;
- впервые введено понятие 'лингвистический профиль пользователей скрытого сообщества', под которым понимается набор лингвистических коррелятов, характеризующих особенности построения пользовательских постов с общим тематическим компонентом.

### **Практическая значимость результатов проведенного исследования**

Практическая значимость исследования состоит в том, что разработанную методику можно использовать в современных социальных сетях с целью создания автоматической системы модерации групп пользователей, которая учитывает их предпочтения по ряду лингвистических параметров.

### **Ценность научных работ соискателя**

Ценность научных работ соискателя заключается в том, что разработка принципов и методов создания и описания скрытых сообществ вносит вклад в развитие корпусной лингвистики и использование корпусов для решения задач прикладной лингвистики. Опубликованные работы имеют значение для современной корпусной лингвистики, способствуя дальнейшей разработке проблем лингвистического профилирования.

### **Полнота изложения материалов диссертации в работах, опубликованных соискателем**

Результаты изложены в 7 научных статьях, 3 из которых опубликованы в рецензируемых научных изданиях, входящих в перечень, утвержденный Высшей аттестационной комиссией при Министерстве науки и высшего образования Российской Федерации:

*Статьи в рецензированных научных журналах, рекомендованных ВАК при Министерстве науки и высшего образования РФ:*

1. Мамаев И. Д. Лингвистические параметры для идентификации скрытых сетевых сообществ / И. Д. Мамаев, О. А. Митрофанова // *Terra Linguistica*. – 2024. — Т. 15. — №. 1. — С. 102-115. (0,96/0,67 п.л)
2. Мамаев И. Д. Лингвистические профили скрытых сообществ: морфосинтаксический аспект / И. Д. Мамаев // *Филологические науки. Вопросы теории и практики*. — 2024. — Т. 17. — Вып. 4. — С. 1155-1162. (0,57 п.л)
3. Мамаев И. Д. Кластерный анализ лингвистических профилей скрытых сообществ / И. Д. Мамаев // *Филологические науки. Вопросы теории и практики*. — 2024. — Т. 17. — Вып. 5. — С. 1739-1747. (0,54 п.л)

*Статьи в других научных изданиях:*

1. Mamaev I. Automatic Detection of Hidden Communities in the Texts of Russian Social Network Corpus / I. Mamaev, O. Mitrofanova // *Artificial Intelligence and Natural Language. AINL 2020. Communications in Computer and Information Science*. — Vol. 1292. — Springer, Cham, 2020. — P. 17-33. (0,88/0,61 п.л)
2. Mamaev I. Hidden Communities in the Russian Social Network Corpus: a Comparative Study of Detection Methods / I. Mamaev, O. Mitrofanova // *Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020) co-located with 16th International Conference on Computational and Cognitive Linguistics (TEL 2020)*. — 2020. — P. 69-78. (0,5/0,35 п.л)
3. Mamaev I. Adaptation of Static and Contextualized Topic Modeling Techniques to Hidden Community Detection / I. Mamaev, O. Mitrofanova // *International Conference on Internet and Modern Society*. — Cham: Springer Nature Switzerland, 2022. — P. 85-97. (0,85/0,59 п.л)
4. Мамаев И. Д. Лингвистические особенности обработки текстов социальных сетей при построении модели скрытых сообществ / И. Д. Мамаев // *Инновационные технологии и технические средства специального назначения: Труды четырнадцатой общероссийской научно-практической конференции*. — Т. 2. — 2022. — С. 312-315. (0,35 п.л)

По итогам обсуждения принято следующее заключение.

По актуальности темы и решаемых задач, новизне полученных результатов, объему, теоретической и научно-практической значимости, полноте опубликованных результатов диссертационная работа Мамаева Ивана Дмитриевича «Лингвистическое исследование скрытых сообществ в корпусе социальных медиа с применением мультимодальных тематических моделей» соответствует паспорту научной специальности 5.9.8 Теоретическая, прикладная и сравнительно-сопоставительная лингвистика (филологические науки) согласно номенклатуре научных специальностей, по которым присуждаются ученые степени (Приказ Министерства науки и высшего образования Российской Федерации от 24.02.2021 № 118).

Диссертация соответствует требованиям, установленным пунктами 9-14 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства РФ от 24.09.2013 № 842 (с изменениями и дополнениями) «О порядке присуждения ученых степеней». В диссертации отсутствуют недостоверные сведения об опубликованных соискателем ученой степени работах.

Диссертация И. Д. Мамаева «Лингвистическое исследование скрытых сообществ в корпусе социальных медиа с применением мультимодальных тематических моделей» рекомендуется к защите на соискание ученой степени кандидата филологических наук

по специальности 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика (филологические науки).

Диссертация прошла обсуждение на кафедре математической лингвистики Санкт-Петербургского государственного университета «21» мая 2024 г. (протокол 44/16/8-02-2).

Проект заключения принят экспертной группой, сформированной распоряжением директора Центра экспертиз М. А. Ревазова от 24.06.2024 № 3131/1р в составе 3 человек: д. филол. наук, профессор Марусенко М.А. (председатель), д. филол. н., профессор, профессор кафедры математической лингвистики СПбГУ, к. филол. н., Гребенников А.О., к. филол. н., доцент, доцент кафедры математической лингвистики СПбГУ, к. филол. н., Хохлова М.В, к. филол. н., доцент, доцент кафедры математической лингвистики СПбГУ.

Результаты очного голосования: «за» – 3 чел., «против» – нет, «воздержались» – нет, протокол №1 от «3» июля 2024 г.

Председатель экспертной группы  
Марусенко Михаил Александрович,  
доктор филологических наук, профессор,  
профессор кафедры математической лингвистики  
федерального государственного бюджетного  
образовательного учреждения высшего образования  
«Санкт-Петербургский государственный университет»

\_\_\_\_\_ (М. А. Марусенко)

ПОДПИСЬ РУКИ

*Марусенко М.А.*  
УДОСТОВЕРЯЮ

