



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)
ФГАОУ ВО «ЮУрГУ (НИУ)»

ЮУрГУ

Проспект Ленина, 76, Челябинск, Россия 454080, тел./факс (351)267-99-00, e-mail: info@susu.ru, www.susu.ru
ОКПО 02066724, ОГРН 1027403857568, ИНН/КПП 7453019764/745301001

№ _____
На № _____ от _____

УТВЕРЖДАЮ

Первый проректор-проректор
по научной работе

ФГАОУ ВО "Южно-Уральский

государственный университет

(национальный исследовательский
университет)"

доктор технических наук, доцент

А.В. Коржов

* « 21 » 10 2024 г.



ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

федерального государственного автономного образовательного учреждения
высшего образования «Южно-Уральский государственный университет

(национальный исследовательский университет)»

о диссертации Мамаева Ивана Дмитриевича

на тему «Лингвистическое исследование скрытых сообществ в корпусе
социальных медиа с применением мультимодальных тематических моделей»,
представленной на соискание ученой степени кандидата филологических
наук по специальности 5.9.8. Теоретическая, прикладная и сравнительно-
сопоставительная лингвистика

Диссертационное исследование И.Д. Мамаева посвящено разработке методики выявления скрытых сетевых сообществ на основе тематического моделирования и последующему лингвистическому профилированию сообществ на основе анализа морфологических, синтаксических и лексических параметров сообщений пользователей с применением аппарата математической статистики.

Актуальность данного исследования обусловлена необходимостью глубокого понимания динамики взаимодействия в этих сообществах. Анализ сообществ может помочь в выявлении суицидальных наклонностей или криминальных и экстремистских ячеек, что требует внимания со стороны правоохранительных органов. Изучение моделей коммуникации с помощью

лингвистических профилей может внести значительный вклад в понимание образа мысли участников и их восприятия окружающего мира, позволит выявлять потенциальные угрозы через распознавание планирования незаконных действий и анализ поведенческих моделей пользователей, способствовать выявлению социальных динамик и трендов, что позволит лучше понять, как формируются мнения и как они могут изменяться в зависимости от контекста.

Научная новизна исследования состоит в разработке процедуры выявления скрытых сообществ посредством применения методов машинного обучения для тематического моделирования сообщений пользователей социальной сети, при этом предложенная процедура тематического моделирования включает выполнение распределения тем по дополнительному параметру – авторам сообщений; определены критерии для сбора корпуса текстов для моделирования выявления скрытых сообществ; определена процедура построения лингвистических профилей скрытых сообществ на основе корреляционного анализа по заданным автором диссертации морфологическим, синтаксическим и лексическим параметрам.

Теоретическая значимость данного исследования заключается в том, что его результаты вносят вклад в развитие компьютерной и математической лингвистики, так как в работе предлагается методологический аппарат, опирающийся на методы машинного обучения и математической статистики, для исследования речевой деятельности пользователей сети, который способствует более глубокому пониманию функционирования языка в скрытых сообществах, позволяет выявить закономерности в языковых практиках участников этих групп. Работа имеет междисциплинарный потенциал, поскольку внедрение лингвистического анализа в существующие методики выделения групп пользователей вносит вклад в исследования скрытых сообществ в социологии, психологии, антропологии, криминологии и других областях.

Практическая значимость работы состоит в том, что предложенные методики могут стать основой для проведения исследований и практической деятельности в сфере социальной информатики, социологии, политологии, психологии и других смежных областях. Разработанные методики могут быть внедрены в сервисы социальных сетей и использоваться для модерации сообществ, мониторинга общественного мнения и прочих задач, требующих извлечения и анализа контента, создаваемого пользователями сети.

Достоверность научных положений, выводов и практических результатов автора подтверждается исходными теоретико-методологическими позициями, предусматривающими довольно широкое рассмотрение положений в области компьютерной и корпусной лингвистики, библиографический аппарат насчитывает 179 наименований, в том числе 95 на иностранных языках; репрезентативностью выборки, используемой как материал исследования, что обусловлено системным подходом к отбору корпуса текстов на основе сформулированных критериев (включены сообщения объемом более 200 символов ограниченного временного периода, опубликованных сбалансированной по полу выборкой из 704 пользователей сети); корректным применением аппарата тематического моделирования и математической статистики для обработки материала исследования, что подтверждается

приведенными выкладками; адекватным задачам этапов исследования применением широкого набора программно-инструментальных средств моделирования; иллюстративным материалом, наглядно подтверждающим выводы автора.

В **первой главе** «Веб как корпус: лингвистические особенности функционирования интернет текста» посвящена анализу лингвистических особенностей интернет-текстов, рассматриваемых в контексте их функционирования в виртуальном пространстве.

Во **второй главе** «Теоретические основания междисциплинарных исследований скрытых сообществ» автор вводит понятие скрытого сообщества. Также автор дает представление о собственной дискретной автор-тематической модели, которая позволяет на основе тематических микромоделей каждого автора с полученными вручную метками тем построить тематическую модель корпуса.

В **третьей главе** «Разработка процедуры лингвистического профилирования скрытых сообществ» автор приводит блок-схему экспериментальной части работы и последовательно показывает реализацию этапов эмпирического исследования. Значительная часть третьей главы посвящена построению и описанию лингвистических профилей полученных сообществ по авторской методике.

К **достоинствам** выполненной работы следует отнести подробный обзор исследований скрытых сообществ для обнаружения групп, демонстрирующих асоциальное и антисоциальное поведение, маргинальных групп, спамеров и т.п. Представлен вполне репрезентативный практический материал исследования, позволяющий подтвердить выдвинутую автором гипотезу.

Методология исследования достаточно продуманная и консистентная. Диссертант прибегает к разнообразным исследовательским методам и программному инструментарию. Автор последовательно решает поставленные задачи, подробно, с применением визуальных средств иллюстрирования излагает результаты каждого этапа своей работы.

Представляют несомненный научный интерес предложенная процедура дискретного автор-тематического моделирования как инструмента выявления скрытых сообществ, а также авторская методика проведения лингвистического профилирования через анализ корреляций признаков морфологического, синтаксического и лексического уровней.

Представленная работа, безусловно, вызывает интерес с точки зрения как постановки задачи, так и предлагаемой автором методологии ее решения. Вместе с тем, высказанные положения диссертации вызвали ряд вопросов и замечаний:

1. В положении 8 автор утверждает, что отсутствие статистически значимых результатов по лексическим коррелятам указывает на «лексическую гомогенность постов социальной сети». При этом лексический параметр профилирования включает оценки корреляции между лексическим разнообразием текста и степенью дистантизации. На наш взгляд, автор не вполне убедительно продемонстрировал в своей работе связь между оцениваемой корреляцией по указанным параметрам и выводом о «гомогенности» текстов постов.

2. Нигде на страницах работы мы не нашли четкого определения термина *мультимодальность*, заявленного в теме исследования. При этом, например, на с. 24 автор использует этот термин при описании мемов, что, видимо, подразумевают трактовку мультимодальности как свойства сообщения передавать смысл через множественные каналы (аудиальный, визуальный и т.д.). В положении 3 автор указывает, что «ручное внедрение параметра «автор» в процедуру тематического моделирования позволяет, во-первых, сделать ее мультимодальной...», что, как нам представляется, предполагает иную трактовку модуса. Требуется уточнить, что автор понимает под мультимодальностью.

3. Автор указывает, что модель скрытых сообществ определяется как семантическая сеть, которую, в свою очередь, «можно охарактеризовать как формальное представление знаний, приобретающее вид ориентированного графа» (с. 28). При этом на рис. 2 «Графическое представление модели скрытых сообществ» приведен граф с неориентированными дугами. По мнению автора, какой граф – ориентированный или неориентированный – представляет собой адекватную модель семантической сети для выявления скрытых сообществ?

4. Имеются замечания по терминологическому аппарату. Автором не всегда соблюдается единство используемой терминологии («лексическая насыщенность» vs «лексическое разнообразие»). Также, под «лексической плотностью» обычно понимается доля всех знаменательных слов в тексте¹. Полагаем, что понятие «отношение количества слов *определенной* части речи ко всем словам исследуемого текста» (с. 70, табл. 5) требует обозначения другим термином.

Указанные замечания носят отчасти дискуссионный характер и не снижают теоретической и практической значимости проведенного исследования. Работа имеет практическую направленность и потенциал к развитию. Так, например, может представлять интерес определение меры скрытности сообществ².

Диссертация Мамаева Ивана Дмитриевича на тему «Лингвистическое исследование скрытых сообществ в корпусе социальных медиа с применением мультимодальных тематических моделей» на соискание ученой степени кандидата филологических наук является научно-квалификационной работой, представляющей собой самостоятельное завершённое исследование с высокой степенью обоснованности научных положений, достоверности и новизны полученных выводов, содержание автореферата раскрывает основное содержание глав и разделов диссертации, результаты исследования в полной мере отражены в семи публикациях, в том числе трех публикациях в научных журналах, входящих в Перечень рецензируемых научных изданий Высшей аттестационной комиссии Министерства науки и высшего образования, что полностью соответствует критериям пп. 9-14 «Положения о присуждении ученых степеней», утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, а ее автор, Мамаев Иван Дмитриевич, заслуживает присуждения ученой степени кандидата филологических наук по

¹ См., например, Ure, J. (1971). Lexical density and register differentiation, in Perren, J. E. and Trim, J. L. M. (eds.), *Applications of linguistics*, Cambridge University Press, London, pages 443-452; Eggins, S. (2004). *An Introduction to Systemic Functional Linguistics* (2nd ed.). London, UK: Pinter Pub Ltd. 360 p.

² He, K., Li, Y., Soundardjan, S., Hopcroft, J.E. (2018) Hidden Community Detection in Social Networks. *Information Sciences*. Volume 425, January 2018, pages 92-106.

специальности 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика.

Настоящий отзыв подготовлен научно-педагогическими работниками ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)»:

1. заведующим кафедрой лингвистики и перевода, кандидатом филологических наук (научная специальность 10.02.21 «Прикладная и математическая лингвистика»), доцентом Бабиной Ольгой Ивановной;

2. профессором кафедры лингвистики и перевода, доктором филологических наук (научная специальность 10.02.21 «Прикладная и математическая лингвистика»), доцентом Шереметьевой Светланой Олеговной.

Отзыв обсужден и одобрен на заседании кафедры лингвистики и перевода Института лингвистики и международных коммуникаций ЮУрГУ 14 октября 2024 г. (протокол № 3).

Заведующий кафедрой лингвистики
и перевода Института лингвистики
и международных коммуникаций
Южно-Уральского
государственного университета,
кандидат филологических наук, доцент
Бабина Ольга Ивановна

Директор Института лингвистики
и международных коммуникаций
Южно-Уральского
государственного университета,
кандидат педагогических наук, доцент
Ярослава Елена Николаевна

Федеральное государственное автономное образовательное учреждение высшего образования «Южно-Уральский государственный университет (национальный исследовательский университет)».

Почтовый адрес: 454080, Россия, г. Челябинск, просп. Ленина, 76.

Тел.: +7 (351) 267-99-00; e-mail: info@susu.ru; Сайт <https://www.susu.ru>

Подписи работников
установлено

Начальник управления по работе
с кадрами Южно-Уральского
государственного университета

