

МЕЖДУНАРОДНЫЕ, ГЛОБАЛЬНЫЕ И РЕГИОНАЛЬНЫЕ ПРОЦЕССЫ

Международные отношения, глобальные и региональные исследования

INTERNATIONAL, GLOBAL AND REGIONAL PROCESSES

International Relations, Global and Regional Studies

Научная статья

УДК 303

[https://doi.org/10.53658/RW2026-4-1\(19\)-24-42](https://doi.org/10.53658/RW2026-4-1(19)-24-42)

Политические науки

Тематическое моделирование в исследованиях цифровой дипломатии: обзор практик и методические рекомендации

Александра Константиновна Бордунос^{1a}, Софья
Владимировна Кошелева^{2b}, Сергей Юрьевич
Севрюков^{3c}, Наталья Владимировна Волкова^{4d}

^{1,2,3} Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

⁴ НИУ Высшая школа экономики, Санкт-Петербург, Россия

^a a.bordunos@gsom.spbu.ru, <https://orcid.org/0000-0003-0347-3180>

^b kosheleva@gsom.spbu.ru, <https://orcid.org/0000-0002-7767-5917>

^c s.sevryukov@spbu.ru, <https://orcid.org/0000-0003-0143-9966>

^d nv.volkova@hse.ru, <https://orcid.org/0000-0002-9045-4393>

Аннотация. В современном контексте цифровая дипломатия претерпевает стремительную трансформацию, обусловленную как внутренними вызовами, так и внешними факторами, такими как фрагментация Интернета и деплатформинг. Тематическое моделирование становится важным методом для анализа этих процессов, позволяя исследовать большой объем неструктурированных текстовых данных, выявляя ключевые тематические паттерны и скрытые смысловые структуры в этой сфере. Цель настоящего обзора – систематизировать и стандартизировать практики применения тематического моделирования в исследованиях цифровой дипломатии на основе анализа опубликованных эмпирических работ. Исследование

основано на систематическом обзоре литературы с использованием протокола PRISMA и поиском в базах данных Scopus и Open Alex ($k = 273$). Обзор структурирован в соответствии с четырьмя этапами тематического моделирования, объединенными в акроним «ТЕМА»: 1) тексты, 2) единообразие (предобработка и очистка данных), 3) модель и 4) анализ. Систематизация практик по этим этапам позволяет предложить исследователям методические рекомендации для всестороннего и обоснованного анализа данных в области цифровой дипломатии. В рамках работы было сформулировано двенадцать исследовательских вопросов: по три для каждого этапа тематического моделирования. Эти вопросы касаются обоснования релевантности и репрезентативности выборки (RQ 1-3), методов очистки текста и работы с нестандартными элементами (RQ 4-6), аргументации выбора модели и ее параметров (RQ 7-9), а также подходов к интерпретации и визуализации результатов (RQ 10-12). Полученные результаты способствуют формированию стандартов и повышению качества исследований в данной междисциплинарной области.

Ключевые слова: тематическое моделирование, цифровая дипломатия, обработка естественного языка, текст-майнинг, латентное размещение Дирихле

Для цитирования: Бордунос А.К., Кошелева С.В., Севрюков С.Ю., Волкова Н.В. Тематическое моделирование в исследованиях цифровой дипломатии: обзор практик и методические рекомендации // Россия и мир: научный диалог. 2026. № 1(19). С. 24-42, [https://doi.org/10.53658/RW2026-4-1\(19\)-24-42](https://doi.org/10.53658/RW2026-4-1(19)-24-42)

Original article

[https://doi.org/10.53658/RW2026-4-1\(19\)-24-42](https://doi.org/10.53658/RW2026-4-1(19)-24-42)

Political Sciences

Topic Modeling in Digital Diplomacy Research: Review of Practices and Methodological Recommendations

Aleksandra K. Bordunos^{1a}, Sofia V. Kosheleva^{2b}, Sergey Yu.
Sevryukov^{3c}, Natalia V. Volkova^{4d}

^{1,2,3} Saint-Petersburg State University, Saint-Petersburg, Russia

⁴ HSE University, Saint-Petersburg, Russia

^a a.bordunos@gsom.spbu.ru, <https://orcid.org/0000-0003-0347-3180>

^b kosheleva@gsom.spbu.ru, <https://orcid.org/0000-0002-7767-5917>

^c s.sevryukov@spbu.ru, <https://orcid.org/0000-0003-0143-9966>

^d nv.volkova@hse.ru, <https://orcid.org/0000-0002-9045-4393>

Abstract. In the modern context, digital diplomacy is undergoing rapid transformation, driven by both internal challenges and external factors such as internet fragmentation and deplatforming. Topic modeling has become an important method for analyzing these processes, allowing researchers to analyze a large volume of unstructured textual data, identify key thematic patterns and uncover hidden semantic structures in the field. The aim of this review is to systematize and standardize the practices in applying

topic modeling to digital diplomacy research based on an analysis of published empirical studies. The study was conducted using a systematic literature review following the PRISMA protocol, searching in Scopus and the Open Alex data bases (k = 273). The review was structured according to four stages of topic modelling: 1) text, 2) data preprocessing and cleansing, 3) modelling, and 4) analysis. The systematization of practices across these stages allows us to offer methodological recommendations for a comprehensive and well-grounded analysis of data in the field of digital diplomacy. As part of the work, twelve research questions were formulated: three for each stage of topic modeling. These questions concern the justification of the relevance and representativeness of the sample (RQ 1-3), text cleaning methods and handling of non-standard elements (RQ 4-6), the rationale for model selection and its parameters (RQ 7-9), as well as approaches to the interpretation and visualization of results (RQ 10-12). The results obtained contribute to the formation of standards and the improvement of research quality in this interdisciplinary field.

Keywords: thematic modeling, digital diplomacy, NLP, LDA

For citation: Bordunos A.K., Kosheleva S.V., Sevryukov S.Yu., Volkova N.V. Topic Modeling in Digital Diplomacy Research: Review of Practices and Methodological Recommendations. *Russia & World: Scientific Dialogue*. 2026; 1(19): 24-42, [https://doi.org/10.53658/RW2026-4-1\(19\)-24-42](https://doi.org/10.53658/RW2026-4-1(19)-24-42)

Введение

Цифровая дипломатия – это набор методов и инструментов [4] для продвижения государственных интересов на внешней и внутренней арене [2], использующих цифровые технологии и платформы, позволяющие установить диалог с разными аудиториями через Интернет и социальные сети [1]. Специфика исследований в области цифровой дипломатии заключается в высокой динамичности, разнообразии форматов – от текстов и видео до интерактивных публикаций, а также масштабе данных, создаваемых ежедневно (если не ежесекундно) и доступных для анализа. Данные в области цифровой дипломатии часто представляют собой неформальные, краткие и контекстуально зависимые сообщения, дополненные визуальными элементами, что требует их тщательной очистки и подготовки для анализа. Кроме того, большое разнообразие и объем источников данных предъявляют высокие требования к методам анализа, а изменчивость тем в режиме реального времени создает вызовы для интерпретации и устойчивости моделей. Среди методов, которые могут учесть эту специфику, выделяется тематическое моделирование, включая динамическое тематическое моделирование, которое позволяет перейти от кросс-секционного к лонгитюдному анализу.

Тематическое моделирование (topic modeling) – относительно новый метод анализа данных в области дипломатии. Он позволяет автоматически сгруппировать тексты по темам на основе статистического распределения ключевых слов, что помогает систематизировать и интерпретировать мультимедиа-массивы информации. Однако специфика выбранной области исследования подчеркивает важность адаптации тематического моделирования именно под условия цифровой дипломатии, что и стало мотивацией для написания данной работы.

Таким образом, цель исследования – систематизировать практики применения тематического моделирования в области цифровой дипломатии на основе опубликованных эмпирических исследований.

Работа служит методологическим ориентиром для исследователей, заинтересованных в осмысленном применении данного метода, в том числе в гуманитарных и социальных науках. Тематическое моделирование относится к методам обработки естественного языка, которые в технических науках широко развиты и применяются, в частности, для обучения и оптимизации моделей машинного обучения. В гуманитарных и социальных науках этот подход пока не получил такого же распространения и стандартизации, что объясняет отсутствие устоявшихся бенчмарков и методологических ориентиров. При этом именно в этих областях наблюдается стремительный рост объема и разнообразия текстовых данных, требующих современных подходов для системного и масштабного анализа. Актуальность тематического моделирования в гуманитарных и социальных науках определяется его способностью выявлять скрытые смысловые структуры и тенденции в больших и неоднородных корпусах текстов, что позволяет расширить возможности интерпретации сложных социально-культурных явлений. Таким образом, развитие методологии тематического моделирования и ее адаптация именно под специфику гуманитарных и социальных исследований представляют собой важный этап в развитии цифровых методов анализа и способствуют формированию новых стандартов и бенчмарков в данных научных сферах.

Материалы и методы

Для обеспечения комплексного охвата источников для обзора были использованы публикации на двух языках: русском и английском. Поиск источников осуществлялся в двух крупных авторитетных базах данных: Open Alex¹ и SCOPUS². В качестве поисковых запросов использовались ключевые словосочетания «тематическое моделирование» («topic modeling») и «цифровая дипломатия» («digital diplomacy») в названии, аннотации и ключевых словах.

Обзор литературы был проведен с использованием протокола PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)³, который обеспечивает прозрачный и систематический подход к поиску, отбору и анализу научных источников.

На первом шаге из каталога публикаций SCOPUS было выделено 217 источников, которые содержали ключевые слова «topic modeling» и «review». Работы преимущественно представляли собой примеры использования метода «тематическое моделирование» с целью проведения обзора литературы в различных направлениях исследования. Только 11 из них были посвящены обзору самого метода

1 См.: OpenAlex. URL: <https://openalex.org/>.

2 См.: SCOPUS. URL: <https://www.scopus.com/>.

3 См.: PRISMA Statement. URL: <https://www.prisma-statement.org/protocols/>.

применительно к анализу текстов в широком смысле слова. Ни одна из публикаций не затрагивала специфику цифровой дипломатии.

На втором шаге из каталога публикаций SCOPUS было выделено 18 статей по ключевым словам «topic modeling» и «diplomacy». Всего 6 из них содержали прямое упоминание цифровой дипломатии. Среди других встречались такие слова, как «мягкая сила» [9; 22], «культурная дипломатия» [8], «знаниевая дипломатия» [5]. Ввиду малого количества публикаций принято решение расширить выборку за счет поиска в открытом каталоге научных публикаций Open Alex. По ключевым словам «topic modeling» и «цифровая дипломатия» отобрано 38 статей из реферируемых научных журналов. После скрининга аннотаций количество релевантных публикаций сократилось до 14, а последующий глубокий анализ полного текста позволил выделить 9 работ, наиболее соответствующих цели исследования. Отбор материалов проводился по заранее определенным критериям релевантности, включающим тематику исследования, использование метода тематического моделирования в анализе цифровой дипломатии, а также наличие эмпирических данных (текстов для анализа).

Дальнейший анализ включал систематизацию ответов на ключевые методологические вопросы: объем и характеристики выборки, методы очистки текстовых данных, выбор моделей анализа и методы визуализации результатов. Такой подход соответствует четырем шагам тематического моделирования, которые для удобства можно объединить в акроним «ТЕМА»:

Т – тексты: для цифровой дипломатии корпус текстов могут составить сообщения в социальных сетях, официальные цифровые публикации, дипломатические отчеты и другие виды текстов. Исследователю важно ответить на вопросы о релевантности собранных текстов, полноте отражения интересующих тематических областей, достаточности количества наблюдений. Ключевые исследовательские вопросы соответствующего раздела следующие:

RQ 1: Как обосновывалась релевантность текстов для исследования цифровой дипломатии?

RQ 2: Как обосновывалась репрезентативность и достаточность выборки текстов для проведения тематического моделирования?

RQ 3: Как обосновывалась сбалансированность покрытия ключевых аспектов изучаемой темы?

Е – единообразие: далее проводится приведение данных к единому формату. В случае с сообщениями из социальных сетей в них могут встречаться хештеги, иконки, смайлики – стоит ли их считать шумом, удаляя, или включать в анализ? Ключевые исследовательские вопросы соответствующего раздела следующие:

RQ 4: Какие методы очистки и предобработки текста были выбраны для повышения качества моделирования?

RQ 5: Какой набор стоп-слов использовался и как обращались с мультимодальными или нестандартными элементами текста?

RQ 6: Каким образом обеспечивалось единообразие формата и устранения шума: какие библиотеки, какое программное обеспечение?

М – модель: выбор модели тематического анализа основывается на целях исследования и особенностях данных, а параметры модели (например, количество тем) подбираются с учетом баланса между качеством интерпретации тем и их статистической обоснованностью. На этом шаге могут быть выявлены ошибки, допущенные на предыдущих шагах, что может потребовать возврата к предыдущим шагам. Например, может потребоваться вернуться на шаг назад и дополнить словарь стоп-слов. Ключевые исследовательские вопросы в данном разделе обзора следующие:

RQ 7: Какие модели выбирали и как аргументировали выбор для имеющейся структуры и объема данных; приходилось ли дополнительно повышать стабильность модели и каким образом?

RQ 8: Как определяли оптимальное число тем (параметры модели) для максимальной интерпретируемости и валидности?

RQ 9: Какие метрики использовать для оценки качества и стабильности модели?

А – анализ: на финальном шаге проводится интерпретация полученных тем, включая анализ ключевых слов, распределения тем по документам и их динамику. На этом этапе важно корректно подобрать подходящие методы визуализации, сравнительного анализа и обсуждения значимости тем в контексте цифровой дипломатии, что позволяет выделить основные тренды и смысловые структуры в данных. Ключевые вопросы следующие:

RQ 10: Какие методы визуализации и аналитики были выбраны для представления результатов?

RQ 11: Как представляли результаты в исследовании?

RQ 12: Какими еще методами дополняли исследование?

Представленная в данном разделе методология обзора позволила выявить сложившиеся практики применения тематического моделирования в цифровой дипломатии, отраженные в современных научных публикациях на русском и английском языках.

Результаты исследования

Тексты: формирование корпуса документов

Набор текстов, а также их метаданных, подобранных и обработанных по определенным правилам для последующего анализа, в тематическом моделировании называют «корпусом документов».

RQ 1: Как обосновывалась релевантность текстов для исследования цифровой дипломатии?

Классический подход к обоснованию релевантности выбранных данных, помимо связи с исследовательскими вопросами, – обращение к теоретическим

основаниям. Проведенный обзор позволил выявить несколько подходящих ключевых концепций, каждая из которых поддерживает тематическое моделирование через акцент на смысловых структурах и коммуникационных паттернах.

Прежде всего, цифровая дипломатия рассматривается как проявление теории «мягкой силы» (soft power) [19]. В теории, сформулированной Джозефом Найем [18], выделяют три ключевых источника влияния государства без применения прямого принуждения: привлекательность культуры, политические ценности и внешняя политика. В контексте формирования корпуса документов для тематического моделирования это обосновывает использование текстовых материалов из официальных каналов о культуре исследуемых стран, таких как Корейская служба культуры и информации (KOCIS) и Японская национальная туристическая организация (JNTO) [19]. Также обращаются к подходу Гын Ли [12], связанному с пятью важными аспектами «мягкой силы»: 1) формирование позитивного и вызывающего доверие образа страны; 2) привлечение поддержки иностранных государств в реализации своей политики; 3) изменение восприятия и предпочтений зарубежной аудитории; 4) укрепление внутреннего единства общества; 5) повышение популярности руководства внутри страны. Эти аспекты облегчают не только выбор источников, но также интерпретацию выявленных тем.

Другим важным теоретическим подходом становятся стратегические нарративы – истории, которые связывают события, участников, локации и помогают скорректировать ожидания и отношения целевой аудитории к явлениям [10]. Выделяют три основных типа нарративов [14]: нарративы системы (описание структуры и ключевых акторов мировой политики), ролевые нарративы (ценности и цели политического субъекта) и нарративы проблем (конкретные политические действия и решения конфликтов). Тематическое моделирование здесь позволяет выявить скрытые темы и паттерны в таких нарративах, что особенно полезно для анализа изменений в коммуникациях, например, смены нарративов ЕС в разные периоды – от «хорошего соседа» [17] до «фасилитатора восстановления» во время пандемии COVID-19 [15]. Расширение и углубление анализа – от государств к объединениям стран, городам или даже отдельным личностям – позволяет исследователям обратить внимание на различные типы коммуникаций [19; 6]. Таким образом, теоретические основы цифровой дипломатии могут дополняться концепциями парадипломатии, такими как региональная, муниципальная, городская дипломатия, которые расширяют арсенал аналитических рамок, учитывая роль негосударственных и субнациональных акторов. В этом контексте тематическое моделирование помогает выявлять ключевые темы и дискурсы в сообщениях городов и регионов, которые стремятся формировать международный имидж и влиять на политику [6].

Третий выявленный в процессе обзора исследований подход – фрейминг – представляет собой способ структурирования информации через социально разделяемые «рамки» («фреймы»), формирующие интерпретацию и понимание событий [7]. В тематическом моделировании это позволяет группировать

темы по типам фреймов, например, разделяя на диагностические (выявление проблем), прогностические (предложения решений) и мотивационные (призывы к действиям) [24].

Помимо теоретических оснований релевантность выбранных данных подтверждается легитимностью используемых социальных сетей для выбранного контекста [15] и высокой активностью отобранных каналов. Например, при анализе цифровой дипломатии Южной Кореи и Японии сначала были представлены списки всех каналов официальных организаций, занимающихся цифровой дипломатией. Из них затем были выбраны наиболее активные страницы на основе количества публикаций и вовлеченности пользователей (комментарии, лайки, репосты) [19]. Аналогично в исследовании Марии Базлуцкой [2]: актуальность выбранной платформы Telegram демонстрируется за счет роста ее популярности среди пользователей, а также наличия официальных каналов государственных структур и аффилированных лиц, что подчеркивает ее значение для цифровой дипломатии. Дополнительно релевантность усиливается отсылкой к важности изучения альтернативных платформ, таких как Telegram, которые в исследованиях используются реже, но становятся особенно ценными в условиях ограничений доминирующих сервисов и разделения Интернета на изолированные сегменты (деплатформизация и фрагментация).

RQ 2: Как обосновывалась репрезентативность и достаточность выборки текстов для проведения тематического моделирования?

В проанализированных текстах нет прямого ответа на эти вопросы. Высокую достоверность источников обеспечивает выбор официальных аккаунтов государственных организаций [19]. В исследовании по материалам из Telegram [2] финальная выборка составила все доступные для анализа посты для двух целевых каналов ($k = 4\ 019$ и $k = 1\ 959$). При длительном этапе существования ресурсов, как в случае KOCIS и JNTO [19], точка отсчета для собранных публикаций может быть задана произвольно (с января 2016 г.), а финальная точка – дата сбора информации (по июнь 2017 г.). В указанном примере финальная выборка для анализа составила 16 962 комментария от 6 083 пользователей на странице KOCIS и 8 278 комментариев от 4 236 пользователей на странице JNTO.

RQ3: Как обосновывалась сбалансированность покрытия ключевых аспектов изучаемой темы?

Сбалансированность выборки достигалась интеграцией разных источников и типов данных: помимо постов и комментариев на них, твитов и ретвитов, корпус текстов может включать популярные хештеги [6]. В сравнительном исследовании Такуто Сакамото и др. [21] дополнили твиты транскриптами всех публичных заявлений представителей различных стран и организаций на официальных заседаниях Совета Безопасности, взяв за дату отсчета 2015 г., когда все пять постоянных членов Совета Безопасности (P5) завели свои аккаунты в исследуемой социальной сети.

В ряде исследований сбалансированность определялась выбранными теоретическими основаниями, например, в случае изучения нарративов, связанных с кризисом, важно было выделить тексты, отображающие нарративы в разные периоды времени: предшествующий кризису период, период кризиса, последующий период и повторная волна кризиса [15].

При изучении конфронтационной дипломатии «боевых волков» в Китае на примере риторики представителя китайского МИД Чжао Лицзяня были выбраны две наиболее популярные платформы – анализировались твиты для внешней аудитории и сообщения в Weibo для внутренней [13]. При этом специфика каждой платформы тоже учитывалась, например, Weibo предлагает возможность анализа суперхештегов, или супертег, что невозможно получить в рамках другой платформы. В данном исследовании выбранный период для сообщений аргументируется определенными событиями: начинается с введения первых таможенных санкций США в отношении Китая и заканчивается подписанием соглашения о первой фазе торговой сделки между двумя странами.

Помимо этого, раскрываются инструменты для сбора текстов, источники данных и сами данные, а также шаги для отсева наиболее релевантных текстов.

Таким образом, формирование корпуса документов тщательно аргументировано с точки зрения актуальности, полноты и сбалансированности, что обеспечивает надежную основу для проведения тематического моделирования и глубокого понимания коммуникационных стратегий цифровой дипломатии.

Единообразие: предварительная подготовка текстов

RQ4: Какие методы очистки и предобработки текста были выбраны для повышения качества моделирования?

Наиболее частые указанные в статьях методы очистки и предобработки текста можно представить в виде акронима «ЛАСТИК»: лемматизация (или стемминг), анализ на наличие шума, выделение стоп-слов, токенизация, корректировка регистра. Лемматизация – приведение слова к его словарной форме, а стемминг – выделение корня слова. Шум – элементы, которые не несут полезной информации, например ошибки, случайные символы, лишние пробелы, хештеги, эмодзи, пунктуация, даты, порядковые числа, элементы разметки. Стоп-слова – часто встречающиеся слова, которые не влияют на смысловой анализ, например, служебные части речи. Токенизация – разбиение текста на смысловые единицы (токены), которые становятся базовыми элементами для последующего анализа. Помимо простых слов (униграмм) важной частью токенизации могут быть комбинации из нескольких слов, которые называются «n-граммами». N-граммы – это словосочетания из n-последовательных слов, используемые для сохранения контекстной информации. Например, биграммы (n = 2) – это пары из двух слов, стоящих рядом в тексте, например, «цифровая дипломатия». Такие словосочетания позволяют сохранить смысловые зависимости, которые теряются при разбиении на отдельные слова.

В исследуемых статьях помимо стандартной корректировки регистра все тексты приводились к единому стандарту написания слов, включая перевод британского английского на американский [21], нормализацию аббревиатур (например, «int'l» → «international») [21] и унификацию написания ключевых слов («COVID-19» и его варианты) [21; 15]. В случае многоязычных корпусов применялся машинный перевод с выборочной проверкой корректности [13].

В некоторых случаях отдельные шаги аргументированно пропускались, например, лемматизация не использовалась, поскольку, по мнению авторов, она снижала интерпретируемость результатов анализа [15].

RQ5: Какой набор стоп-слов использовался и как обращались с мультимодальными или нестандартными элементами текста?

Для удаления стоп-слов использовались стандартные наборы из доступных библиотек, таких как spaCy, дополненные характерными для социальных сетей сокращениями и символами: «rt» (retweet), «amp» (&), «re» (reply), а также сокращениями английских грамматических форм («s», «d», «ve» и др.) [15]. К шумовым элементам относили эмодзи (эмодзи с флагами заменялись на текстовые названия стран, остальные удалялись) [21], числовые данные и символы хештегов, если только хештеги не являлись объектом анализа [6].

RQ6: Каким образом обеспечивалось единообразие формата и устранения шума: какие библиотеки, какое программное обеспечение?

Для единообразия формата и устранения шума авторами применялись программные инструменты и библиотеки на базе Python 3, в частности spaCy и scikit-learn (включая CountVectorizer, который использовался для преобразования текстов в матрицу частот встречаемости слов) [15]. В отличие от GloVe, который создает плотные векторные представления слов с учетом их контекста и глобальной статистики, CountVectorizer формирует разреженные матрицы частот, что подходит для простых и быстрых решений, но уступает по точности семантического анализа. Он создает словарь всех уникальных слов в корпусе и для каждого документа считает, сколько раз каждое слово в нем встречается. В результате каждый текст представляется в виде числового вектора с частотами встречаемости слов.

spaCy представляет собой современную и производительную библиотеку для обработки естественного языка (Natural Language Processing, NLP), разработанную на языке Python с элементами Cython. Основное назначение языка программирования Cython – оптимизация производительности вычислительно тяжелых задач и библиотек, интеграция с нативным кодом, а также написание расширений для Python с возможностью гибкого управления памятью и типами данных. Преимущество выбранных библиотек – в возможности работать с текстами на разных языках, включая русский. Однако в настоящее время уже есть возможность подключать библиотеки, которые изначально специализируются на работе с русским языком, например Py morphology, Py system.

Автоматизированные способы предобработки дополнялись пользовательскими скриптами [2] и такими инструментами, как ConText [19], что обеспечивало высокое качество и воспроизводимость подготовки данных к анализу.

Для расширения корпуса применялись эмбединги (представление слов в виде многомерных числовых векторов) и выделение n-грамм. Для этих целей авторы использовали алгоритм GloVe (Global Vectors for Word Representation; глобальные векторы для представления слов) для учета глобальной статистики словоупотребления и выявления семантических связей [21]. Этот алгоритм разработан в Стэнфорде. Его основная идея состоит в извлечении семантических отношений между словами с использованием матрицы совместной встречаемости слов. В отличие от моделей, опирающихся только на выбранные для исследования тексты, GloVe учитывает как глобальную, так и локальную статистику словоупотребления, что позволяет получать осмысленные и точные векторные представления слов⁴. В работе Сакамото и др. [21] эмбединги применяли и к выявлению ближайших по смыслу слов для ключевых понятий («угроза», «защищать», «суверенитет»): выявлялись слова, векторы эмбедингов которых максимально близки к вектору конкретного понятия в пространстве эмбедингов [21]. В рассмотренном примере авторы задавали такие параметры, при которых слова представлялись в виде векторов фиксированной размерности (100 измерений), а размер окна контекста был равным 12 с обеих сторон: при обучении учитывались слова в пределах 12 позиций слева и справа от текущего слова.

Модель: выбор модели тематического анализа

RQ7: Какие модели выбирали и как аргументировали выбор для имеющейся структуры и объема данных; приходилось ли дополнительно повышать стабильность модели и каким образом?

В изучаемых статьях использовались три модели:

- Bertopic (Bidirectional Encoder Representations from Transformers; двунаправленные представления кодировщика из трансформеров);
- LDA (Latent Dirichlet Allocation; латентное распределение Дирихле);
- NMF (Non-negative Matrix Factorisation; неотрицательное матричное разложение).

Для обеспечения стабильности результатов устанавливали параметр random state (обычно 42), что помогало гарантировать воспроизводимость анализа [15]. Параметр random state в машинном обучении задает начальное состояние (seed) генератора псевдослучайных чисел, используемого в различных алгоритмах, что повышает при повторном запуске эксперимента вероятность получения идентичного результата. Если random state не задан или равен None, каждый запуск будет давать разные результаты из-за нового начального

⁴ См.: GloVe: Global Vectors for Word Representation. URL: <https://nlp.stanford.edu/projects/glove/>.

значения генератора. Число 42 стало популярным в связи с книгой «Автостопом по галактике», но само по себе не имеет технического преимущества перед другими значениями.

За рамками исследований цифровой дипломатии осталась дискуссия в отношении уязвимости коротких текстов для анализа с помощью традиционных подходов, даже при сочетании с моделями эмбедингов слов, таких как Word2Vec и GloVe. Модели LDA и NMF применяют метод «мешка слов» и не учитывают порядка слов, что упрощает расчет, но снижает семантическую точность. В обзоре Warsito et al. [23] обратили внимание на интерес к таким специальным моделям (Biterm Topic Modeling – BTM, Dirichlet Multinomial Mixture – DMM), которые анализируют не отдельные сообщения, а весь массив сразу и обнаруживают схожие слова по всей коллекции сообщений. Алгоритм Bertopic выделяется способностью обработки текста с учетом двунаправленного контекста слова, что позволяет глубже и точнее улавливать смысл и взаимосвязи между словами в предложении. Текст сначала переводится в эмбединги, а потом для поиска тем используются алгоритмы кластеризации (например, HDBSCAN). Вдобавок обзор подчеркнул наличие новых разработок, таких как гибридные методы: позиционно-чувствительная тематическая модель (Position-Sensitive Word Embedding Topic Model – PS-WETM), интеграция больших языковых моделей в процесс анализа тем, а также метод семантического компонентного анализа (Semantic Component Analysis – SCA), или Word Embedding Topic Model (WETM), – первый умеет находить не только стандартные темы, но и разные смысловые паттерны в текстах, а второй (WETM) улучшает интерпретируемость за счет вложений слов.

RQ8: Как определяли оптимальное число тем (параметры модели) для максимальной интерпретируемости и валидности?

Есть модели, которые сами определяют количество тем, и те, которые предполагают задание количества тем исследователем. При этом, как описывают авторы, число тем, предлагаемое моделью по умолчанию, могло значительно превышать ожидания и затруднять интерпретацию, например, $k = 260$ [15]. В связи с этим оптимальное количество тем в исследуемых работах определялось эмпирическим путем, основываясь на интерпретируемости и валидности результатов анализа. Исследователи проводили итеративный пробный анализ с числом тем в диапазоне от 10 до 60, выбирая конечное количество тем в пределах 18–30 в зависимости от качества кластеров и информативности. Такой подход позволял найти баланс между детальными и обобщенными тематическими моделями и повысить читаемость результатов [15; 21].

RQ9: Какие метрики использовать для оценки качества и стабильности модели?

Для оценки качества тематической модели в изучаемых статьях использовалась перплексия (perplexity), которая отражает степень соответствия модели данным. Она вычисляется как обратная вероятность тестового набора слов, нормализованная по

количеству слов, и служит мерой «удивления» модели полученными данными: чем ниже значение перплексии, тем лучше модель предсказывает данные и тем точнее она определяет структуру текстов. Сравнение значений перплексии для моделей с разным числом тем помогает определить их оптимальное количество.

Также упоминался подход динамического тематического моделирования, что позволяет отслеживать изменение частоты тем во времени, например, поквартально или ежемесячно. Это помогало подтвердить валидность (face validity) и адекватность модели в контексте реальных событий и изменений дискурса [15].

Альтернативные подходы, которые остались за рамками проанализированных текстов, приведены в обзоре Хоссейни Марани (Hosseiny Marani), Баумер (Baumer) [11]. Первая группа параметров измеряет стабильность модели. Для ее оценки используют совпадение тем или n-грамм в темах, их вероятностное распределение при повторных запусках модели: коэффициент Жаккара (Jaccard similarity), коэффициент Дайса (Dice's coefficient), ранговая корреляция Спирмена, косинусное сходство, дивергенция Кульбака–Лейблера (Kullback-Leibler (KL) divergence). Альтернативный вариант – сравнивать распределение документов по темам при разных запусках одинаковых моделей. Вторая группа параметров позволяет улучшить стабильность модели. Например, эволюционный алгоритм (Differential Evolution – DE) позволяет запускать несколько вариантов моделей с разным количеством тем до момента достижения стабильности результата. При этом сами авторы делают акцент на том, что стабильность не всегда подтверждает качество исследования, поэтому выбор остается за исследователем.

В контексте оценки содержательного качества тем (а не только их статистической адекватности) методологическое сообщество выделяет также метрики семантической когерентности (coherence metrics), такие как C_v или C_umass (см., например: [20]). Данные метрики количественно оценивают интерпретируемость темы на основе анализа совместной встречаемости ключевых слов в исходном корпусе, тем самым приближая оценку к человеческому восприятию связности темы [16].

В проанализированном массиве исследований цифровой дипломатии подобные метрики не получили заметного распространения, что может указывать на известный методический разрыв между чисто алгоритмической оптимизацией и задачами содержательной интерпретации. В качестве рекомендации представляется целесообразным комбинировать несколько подходов: использование перплексии для сравнительной настройки модели, оценку семантической когерентности для отбора наиболее осмысленных тематических наборов и, наконец, обязательную экспертную верификацию (face validity) через соотнесение динамики тем с историческим контекстом, что и наблюдается в наиболее релевантных практиках [15]. Такой многоуровневый подход позволяет преодолеть ограничения отдельных метрик и повысить обоснованность выводов в исследованиях, где тематическая модель служит инструментом для работы со смыслами, а не самоцелью.

Анализ: визуализация результатов, их объяснение авторами

RQ10: Какие методы визуализации и аналитики были выбраны для представления результатов?

Визуализация результатов тематического моделирования чаще всего представляется через таблицы, в которых указываются номер темы, набор ключевых слов и исследовательское название либо краткое описание темы.

Визуальные методы не описываются, хотя для выполнения этого шага существуют специализированные библиотеки для интерактивной визуализации, такие как ruLDAvis для LDA, позволяющие отображать распределение слов и документов по темам, а также динамику изменения тем со временем. RuLDAvis при визуализации представляет темы в виде окружностей, размер которых пропорционален их значимости в корпусе. Пользователь может изучать ключевые слова, регулируя уровень детализации, а также наблюдать вклад слов в различные темы. Такой вариант визуализации удобен для интерпретации, но его трудно перенести в рукопись.

Также есть возможность дополнительно включить визуализацию в стандартных графических библиотеках (matplotlib, seaborn) для построения гистограмм и распределения тем по метаданным (жанрам, времени публикации). В некоторых работах есть потенциал для применения картографической визуализации с использованием GIS-систем для пространственного анализа текстов. В исследованных публикациях этот арсенал не использовался авторами.

RQ11: Как представляли результаты в исследовании?

По аналогии с визуализацией результатов для интерпретации, для представления результатов используют табличный вид, дополненный текстовыми пояснениями с опорой на теоретическую рамку и контекст.

Для больших наборов тем демонстрируются только наиболее популярные кластеры (6 из 18 у Морал (Moral) [15]), а остальные переносятся в приложения или дополнительные материалы, размещаемые не только на сайте издательства, но и на открытых хранилищах, таких как GitHub. Схожим образом презентовали результаты при тестировании разных моделей – в тексте работы приводили таблицу с результатами применения выбранной модели, альтернативные опции предлагали в приложении или дополнительных материалах.

RQ12: Какими еще методами дополняли исследование?

Для углубленного анализа результатов часто использовался метод семантического сетевого анализа (SNA), который позволяет анализировать количество и взаимосвязи слов с учетом их семантических отношений. Этот метод выявляет скрытые связи между словами и темами, расширяя возможности традиционного тематического анализа.

Также применялся анализ пользовательских профилей (user profile analysis), что позволяло учитывать особенности аудитории и контекст коммуникаций [19]. Таким образом, комплексный подход к визуализации и аналитике включал комбинирование табличных, графических и сетевых методов, обеспечивая всестороннее представление и глубокое понимание результатов тематического моделирования.

Выводы

Тематическое моделирование – еще относительно новый инструмент для изучения цифровой дипломатии, что отражает выборка исследований для настоящего обзора. В дальнейшем выборку можно увеличить как за счет добавления синонимов и ключевых слов, связанных с методом («тематический анализ», «LDA», «Bert», «Bertopic»), так и с помощью расширения поля исследования «электронная дипломатия», «интернет-дипломатия», «дипломатия социальных сетей», «городская дипломатия», «Web 2.0 дипломатия», «твипломатия» (Twiplomacy)).

Обзор не только показал пригодность метода «тематическое моделирование» для выбранного поля исследования, но также позволил выявить направление для улучшений процесса анализа. Его применение в исследовательских работах позволяет системно анализировать большие объемы текстовых данных, выделять ключевые темы и проследивать динамику развития тем во времени. Использование современных моделей, таких как Bertopic, LDA и NMF, с применением методов предобработки и визуализации обеспечивает как высокую информативность, так и интерпретируемость результатов.

Среди ограничений тематического моделирования стоит выделить зависимость результатов от качества предобработки и специфики корпуса, что может влиять на интерпретируемость и стабильность модели. Используемые метрики, например перплексия, не всегда полностью отражают смысловую адекватность моделей, требуя дополнительной экспертной оценки. Кроме того, динамическое тематическое моделирование и анализ мультязычных корпусов повышают уровень сложности и могут требовать дополнительных ресурсов и проверок качества.

Важным направлением является разработка стандартизированных протоколов для выбора параметров моделей и оценки их качества с привлечением экспертной оценки. Также перспективно включение методов интерактивной визуализации и аналитики, способствующих более широкому вовлечению исследователей и пользователей в процесс интерпретации результатов.

Список литературы

1. Базлуцкая М.М., Сытник А.Н., Цветкова Н.А. Вовлеченность аудитории как показатель эффективности цифровой дипломатии [Evaluating Audience Engagement as a Measure of Digital Diplomacy Effectiveness] // MGIMO Review of International Relations. 2024. Т. 17. № 4. С. 48–82. <https://doi.org/10.24833/2071-8160-2024-4-97-48-82>.

2. Базлуцкая М.М. Цифровая дипломатия в Telegram: сравнительный анализ вовлеченности аудитории и активности посольств России и США [Digital Diplomacy in Telegram: A Comparative Analysis of Audience Engagement and Activity of the Russian and US Embassies] // Вестник Санкт-Петербургского университета. Международные отношения. 2024. Т. 17. Вып. 4. С. 431–456. <https://doi.org/10.21638/spbu06.2024.403>.
3. Йе Л., Булгарова Б.А., Хань Ч., Цзюй Я. Формирование национального имиджа Китая в кросс-культурном контексте: анализ настроений и тематическое моделирование комментариев к туристическому контенту Bilibili [Formation of China's National Image in a Cross-Cultural Context: Sentiment Analysis and Topic Modeling of Comments on Bilibili Tourism Content] // Litera. 2025. № 5. С. 254–267. <https://doi.org/10.25136/2409-8698.2025.5.74116>. EDN: QMVIF0.
4. Ордовский Д., Кожемякин Е.А. Цифровая публичная дипломатия в эпоху социальных медиа: эволюция и современное состояние [Digital Public Diplomacy in Social Media Era: Evolution and Current Trends] // Полис. Политические исследования. 2025. № 2. С. 157–166. <https://doi.org/10.17976/jpps/2025.02.11>.
5. Adam E., Adam N. Education hubs in Qatar and the United Arab Emirates: Analysis of discursive constructions using Latent Dirichlet Allocation // International Journal of Educational Development. 2024. № 107. <https://doi.org/10.1016/j.ijedudev.2024.103034>.
6. Asdourian B., Chariatte J., Ingenhoff D. Digital City Diplomacy and International Cities Networks: Collaboration and City Branding Strategies around Climate Issues // International Communication Gazette. 2024. № 86(7). P. 541–564.
7. Carter M.J. The Hermeneutics of Frames and Framing: An Examination of the Media's Construction of Reality // SAGE Open. 2013. № 3(2). P. 1–11. <https://doi.org/10.1177/2158244013487915>.
8. Cui Y. Research on the Regional Dissemination Paths of Chinese Culture and Cultural Diplomacy Strategies in the Context of the "Belt and Road Initiative" // Pakistan Journal of Life & Social Sciences. 2025. № 23(1).
9. Gibson J., Istad F. Toward a Political Model of Soft Power: An Empirical Evaluation of South Korean Foreign Policy // Asian Survey. 2025. № 65(3). P. 459–495.
10. Hedling E. Storytelling in EU Public Diplomacy: Reputation Management and Recognition of Success // Place Branding and Public Diplomacy. 2020. № 16(2). P. 143–152. <https://doi.org/10.1057/s41254-019-00138-2>.
11. Hosseiny Marani A., Baumer E.P. A Review of Stability in Topic Modeling: Metrics for Assessing and Techniques for Improving Stability // ACM Computing Surveys. 2023. № 56(5). P. 1–32.
12. Lee G. A Theory of Soft Power and Korea's Soft Power Strategy // Korean Journal of Defense Analysis. 2009. № 21(2). P. 205–218. <https://doi.org/10.1080/10163270902913962>.
13. Li M. A Cross-Platform Comparison of China's Confrontational Diplomatic Communication // The Journal of International Communication. 2024. № 30(2). P. 372–393.
14. Miskimmon A., O'Loughlin B., Roselle L. Forging the World: Strategic Narratives and International Relations. Ann Arbor: University of Michigan Press, 2017. 352 p. <https://doi.org/10.3998/mpub.6504652>.
15. Moral P. Restoring Reputation through Digital Diplomacy: The European Union's Strategic Narratives on Twitter⁵ during the COVID-19 Pandemic // Communication & Society. 2023. P. 241–269.
16. Newman D. et al. Automatic evaluation of topic coherence // Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics. 2010. P. 100–108.
17. Nițoiu C. The Narrative Construction of the European Union in External Relations // Perspectives on European Politics and Society. 2013. № 14(2). P. 240–255. <https://doi.org/10.1080/15705854.2013.785264>.
18. Nye J.S. Soft Power: The Means to Success in World Politics. New York: PublicAffairs Books, 2004. 192 p.
19. Park S., Chung D., Park H.W. Analytical Framework for Evaluating Digital Diplomacy Using Network Analysis and Topic Modeling: Comparing South Korea and Japan // Information Processing & Management. 2019. № 56(4). P. 1468–1483.
20. Röder M., Both A., Hinneburg A. Exploring the space of topic coherence measures // Proceedings of the eighth ACM international conference on Web search and data mining. 2015. P. 399–408.
21. Sakamoto T., Araki M., Ito H., Matsuoka T. How Different are Offline and Online Diplomacy? A Comparative Analysis of Public Statements and SNS Posts by Delegates to the United Nations // Frontiers in big Data. 2024. № 7. P. 1304806.

5 *Доступ к X (бывш. Twitter) на территории Российской Федерации заблокирован на основании требования Генпрокуратуры от 24 февраля 2022 г. в связи с распространением ложной и недостоверной общественно значимой информации, 18+.

22. Tahat K., Mansoori A., Tahat D.N., Habes M., Salloum S. Leveraging Soft Power: A Study of Emirati Online Journalism Through Arabic Topic Modeling // International Conference on Business and Technology. 2023. P. 13–20.
23. Warsito B., Suseno J.E., Arifudin A. Embedding and Topic Modeling Techniques for Short Text Analysis on Social Media: A Systematic Literature Review // Data and Metadata. 2025. № 4. P. 1168–1168.
24. Xiong Y., Cho M., Boatwright B. Hashtag Activism and Message Frames Among Social Movement Organizations: Semantic Network Analysis and Thematic Analysis of Twitter during the #MeToo Movement // Public Relations Review. 2019. № 45(1). P. 10–23.

Информация об авторах

БОРДУНОС Александра Константиновна. Старший преподаватель кафедры организационного поведения и управления персоналом Санкт-Петербургского государственного университета. Руководитель направления ИИ в образовании Центра преподавательского мастерства Санкт-Петербургского государственного университета. <https://orcid.org/0000-0003-0347-3180>. Адрес: Российская Федерация, 199034, г. Санкт-Петербург, Университетская наб., 7–9. a.bordunos@gsom.spbu.ru

КОШЕЛЕВА Софья Владимировна. Доктор психологических наук. Профессор кафедры организационного поведения и управления персоналом Санкт-Петербургского государственного университета. <https://orcid.org/0000-0002-7767-5917>. Адрес: Российская Федерация, 199034, г. Санкт-Петербург, Университетская наб., 7–9. kosheleva@gsom.spbu.ru

СЕВРЮКОВ Сергей Юрьевич. Старший преподаватель кафедры технологии программирования Санкт-Петербургского государственного университета. <https://orcid.org/0000-0003-0143-9966>. Адрес: Российская Федерация, 199034, г. Санкт-Петербург, Университетская наб., 7–9. s.sevryukov@spbu.ru

ВОЛКОВА Наталья Владимировна. Кандидат психологических наук. Доцент НИУ Высшая школа экономики. <https://orcid.org/0000-0002-9045-4393>. Адрес: Российская Федерация, 199034, г. Санкт-Петербург, ул. Почтамтская д. 3-5. nv.volkova@hse.ru

Раскрытие информации о конфликте интересов

Авторы заявляют об отсутствии конфликта интересов.

Информация о статье

Поступила в редакцию: 1 ноября 2025 г. Одобрена после рецензирования: 10 февраля 2026 г. Принята к публикации: 20 февраля 2026 г. Опубликована: 1 марта 2026 г.

Авторы прочитали и одобрили окончательный вариант рукописи.

Информация о рецензировании

«Россия и мир: научный диалог» благодарит анонимных рецензентов за их вклад в рецензирование этой работы.

References

1. Bazlutckaia M.M., Sytnik A.N., Tsvetkova N.A. Evaluating Audience Engagement as a Measure of Digital Diplomacy Effectiveness. Vestnik MGIMO universiteta. Mezhdunarodnye otnosheniya [MGIMO Review of International Relations]. 2024; 17(4):48–82 [In Russian]. <https://doi.org/10.24833/2071-8160-2024-4-97-48-82>.
2. Bazlutckaia M.M. Digital Diplomacy in Telegram: A Comparative Analysis of Audience Engagement and Activity of the Russian and US Embassies. Vestnik Sankt-Peterburgskogo universiteta.

- Mezhdunarodnye otnosheniya [Vestnik of Saint Petersburg University. International Relations]. 2024; 17(4):431–456 [In Russian]. <https://doi.org/10.21638/spbu06.2024.403>.
3. Ye L., Bulgarova B.A., Han Z., Ju Ya. Formation of China's National Image in a Cross-Cultural Context: Sentiment Analysis and Topic Modeling of Comments on Bilibili Tourism Content. Litera. 2025; 5:254–267 [In Russian]. <https://doi.org/10.25136/2409-8698.2025.5.74116>. EDN: QMVIF0.
4. Ordovsky D., Kozhemyakin E.A. Digital Public Diplomacy in Social Media Era: Evolution and Current Trends. Polis. Politicheskie issledovaniya [Polis. Political Studies]. 2025; 2:157–166 [In Russian]. <https://doi.org/10.17976/jpps/2025.02.11>.
5. Adam E., Adam N. Education hubs in Qatar and the United Arab Emirates: Analysis of discursive constructions using Latent Dirichlet Allocation. International Journal of Educational Development. 2024; 107 [In English]. <https://doi.org/10.1016/j.ijedudev.2024.103034>.
6. Asdourian B., Chariatte J., Ingenhoff D. Digital City Diplomacy and International Cities Networks: Collaboration and City Branding Strategies around Climate Issues. International Communication Gazette. 2024; 86(7):541–564 [In English].
7. Carter M.J. The Hermeneutics of Frames and Framing: An Examination of the Media's Construction of Reality. SAGE Open. 2013; 3(2):1–11 [In English]. <https://doi.org/10.1177/2158244013487915>.
8. Cui Y. Research on the Regional Dissemination Paths of Chinese Culture and Cultural Diplomacy Strategies in the Context of the " Belt and Road Initiative ". Pakistan Journal of Life & Social Sciences. 2025; 23(1) [In English].
9. Gibson J., Istad F. Toward a Political Model of Soft Power: An Empirical Evaluation of South Korean Foreign Policy. Asian Survey. 2025; 65(3):459–495 [In English].
10. Hedling E. Storytelling in EU Public Diplomacy: Reputation Management and Recognition of Success. Place Branding and Public Diplomacy. 2020; 16(2):143–152 [In English]. <https://doi.org/10.1057/s41254-019-00138-2>.
11. Hosseiny Marani A., Baumer E.P. A Review of Stability in Topic Modeling: Metrics for Assessing and Techniques for Improving Stability. ACM Computing Surveys. 2023; 56(5):1–32 [In English].
12. Lee G. A Theory of Soft Power and Korea's Soft Power Strategy. Korean Journal of Defense Analysis. 2009; 21(2):205–218 [In English]. <https://doi.org/10.1080/10163270902913962>.
13. Li M. A Cross-Platform Comparison of China's Confrontational Diplomatic Communication. The Journal of International Communication. 2024; 30(2):372–393 [In English].
14. Miskimmon A., O'Loughlin B., Roselle L. Forging the World: Strategic Narratives and International Relations. Ann Arbor: University of Michigan Press, 2017. 352 p. [In English]. <https://doi.org/10.3998/mpub.6504652>.
15. Moral P. Restoring Reputation through Digital Diplomacy: The European Union's Strategic Narratives on Twitter*⁶ during the COVID-19 Pandemic. In Communication & Society. 2023:241–269 [In English].
16. Newman D. et al. Automatic evaluation of topic coherence. Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics. 2010:100–108 [In English].
17. Nițoiu C. The Narrative Construction of the European Union in External Relations. Perspectives on European Politics and Society. 2013; 14(2):240–255 [In English]. <https://doi.org/10.1080/15705854.2013.785264>.
18. Nye J.S. Soft Power: The Means to Success in World Politics. New York: PublicAffairs Books, 2004. 192 p. [In English].
19. S., Chung D., Park H.W. Analytical Framework for Evaluating Digital Diplomacy Using Network Analysis and Topic Modeling: Comparing South Korea and Japan. Information Processing & Management. 2019; 56(4):1468–1483 [In English].
20. Röder M., Both A., Hinneburg A. Exploring the space of topic coherence measures. Proceedings of the eighth ACM international conference on Web search and data mining. 2015:399–408 [In English].
21. Sakamoto T., Araki M., Ito, H., Matsuoka, T. (). How Different are Offline and Online Diplomacy? A Comparative Analysis of Public Statements and SNS Posts by Delegates to the United Nations. Frontiers in big Data. 2024; 7:1304806 [In English].
22. Tahat K., Mansoori A., Tahat D.N., Habes M., Salloum S. Leveraging Soft Power: A Study of Emirati Online Journalism Through Arabic Topic Modeling. In International Conference on Business and Technology. 2023:13–20 [In English].

⁶ *Доступ к X (бывш. Twitter) на территории Российской Федерации заблокирован на основании требования Генпрокуратуры от 24 февраля 2022 г. в связи с распространением ложной и недостоверной общественно значимой информации, 18+.

23. Warsito B., Suseno J.E., Arifudin A. Embedding and Topic Modeling Techniques for Short Text Analysis on Social Media: A Systematic Literature Review. *Data and Metadata*. 2025; 4:1168–1168 [In English].
24. Xiong Y., Cho M., Boatwright B. Hashtag Activism and Message Frames Among Social Movement Organizations: Semantic Network Analysis and Thematic Analysis of Twitter during the #MeToo Movement. *Public Relations Review*. 2019; 45(1):10–23 [In English].

About the authors

- K. BORDUNOS. Senior Lecturer of Department of Organizational Behavior and Personnel Management at St.Petersburg University. Head of AI in Education of Teaching Excellence Lab at St.Petersburg University. <https://orcid.org/0000-0003-0347-3180>. Address: 7–9, Universitetskaya nab., St.Petersburg, 199034, Russian Federation. a.bordunos@gsom.spbu.ru
- Sofia V. KOSHELEVA. DSc. (Psychology). Professor of Department of Organizational Behavior and Personnel Management at St.Petersburg University. <https://orcid.org/0000-0002-7767-5917>. Address: 7–9, Universitetskaya nab., St.Petersburg, 199034, Russian Federation. kosheleva@gsom.spbu.ru
- Sergey Yu. SEVRYUKOV. Senior Lecturer of Department of Programming Technology at St.Petersburg University. <https://orcid.org/0000-0003-0143-9966> Address: 7–9, Universitetskaya nab., St.Petersburg, 199034, Russian Federation. s.sevryukov@spbu.ru
- Natalia V. VOLKOVA. CandSc. (Psychology). Associate Professor of HSE University in St.Petersburg. <https://orcid.org/0000-0002-9045-4393>. Address: 3-5, Pochtamskaya st., St.Petersburg, 199034, Russian Federation. nv.volkova@hse.ru

Contribution of the authors

The authors declare no conflicts of interests.

Article info

Received: November 1, 2025. Approved after review: February 10, 2026. Accepted for publication: February 20, 2026. Published: March 1, 2026.

The authors have read and approved the final manuscript.

Peer review info

«Russia & World: Scientific Dialogue» thanks the anonymous reviewers for their contribution to the peer review of this work.