

Возможности и ограничения использования  
больших текстовых данных  
в отечественной и зарубежной политической науке  
на примере базы данных GDELТ

Денис О. Вакарчук

*Российский государственный гуманитарный университет,  
Москва, Россия, vakarchuk.d@rggu.ru*

*Аннотация.* В статье в сопоставительном ключе рассмотрена практика количественного анализа больших текстовых данных в российской и зарубежной политической науке на примере использования базы данных GDELТ (Global Database of Events, Language, and Tone). На основе обзора научных публикаций отечественных и зарубежных исследователей показано разнообразие исследовательских задач, которые позволяют решить данные GDELТ. При этом задачи, решаемые отечественными политологами посредством данных GDELТ, носят преимущественно описательный характер, тогда как для работ зарубежных ученых характерна инкорпорация данных GDELТ в сложные математические модели с целью определения характера и направления связей между переменными. Выявлено, что исследовательская вариативность обусловлена широким функционалом, которым обладает база GDELТ. К основным преимуществам использования GDELТ автор относит, во-первых, возможность проведения автоматизированного ивент- и контент-анализа больших текстовых данных, во-вторых, дополнительную опцию для операционализации и валидизации политических явлений и процессов, в-третьих, возможность для фальсификации результатов проведенного анализа. В качестве ограничений работы с базой определены, во-первых, несовершенство онтологии CAMEO, во-вторых, неполнота данных, содержащихся в базе, в-третьих, технические ограничения работы с базой GDELТ. В связи с этим автором сделан вывод о необходимости крайне осторожного применения данных GDELТ, тестируя в каждом конкретном случае эти измерения на валидность.

*Ключевые слова:* GDELТ, ивент-анализ, количественные методы, методология политологического анализа

*Для цитирования:* Вакарчук Д.О. Возможности и ограничения использования больших текстовых данных в отечественной и зарубежной политической науке на примере базы данных GDELТ // Вестник РГГУ. Серия «Политология. История. Международные отношения». 2024. № 3. С. 30–46. DOI: 10.28995/2073-6339-2024-3-30-46

## Opportunities and limitations of using big text data in Russian and foreign political science on the example of GDELT database

Denis O. Vakarchuk

*Russian State University for the Humanities, Moscow, Russia,  
vakarchuk.d@rggu.ru*

*Abstract.* In a comparative way the article considers the practice of quantitative analysis of large textual data in Russian and foreign political science using the example of the GDELT (Global Database of Events, Language, and Tone) database. Based on the review of scientific publications by Russian and foreign researchers, the article shows a variety of research tasks that can be solved using GDELT data. At the same time, the tasks solved by national political scientists using GDELT data are mainly descriptive in nature, while the works of foreign scientists are characterized by the incorporation of GDELT data into complex mathematical models in order to determine the nature and direction of relationships between variables. It is revealed that the variability of research is due to the wide functionality of the GDELT database. The main advantages of using GDELT are, firstly, the possibility for automated event and content analysis of large textual data, secondly, an additional option for operationalizing and validating political phenomena and processes, and thirdly, the possibility of falsifying the results of the analysis. The limitations of working with the database were identified as, firstly, the imperfections of the CAMEO ontology, secondly, the incompleteness of the data contained in the database, and thirdly, the technical limitations of working with the GDELT database. Thereby, the author concluded that GDELT data should be used with extreme caution, and that measurements should be checked for validity in each case.

*Keywords:* GDELT, event analysis, quantitative methods, political analysis methodology

*For citation:* Vakarchuk, D.O. (2024) "Opportunities and limitations of using big text data in Russian and foreign political science on the example of GDELT database", *RSUH/RGGU Bulletin. "Political Science. History. International Relations" Series*, no. 3, pp. 30–46, DOI: 10.28995/2073-6339-2024-3-30-46

Несмотря на то что по-прежнему «ознакомление с количественными методами в отечественной международно-политической науке находится на начальном этапе» [Дегтерев 2015, с. 45], в последние годы среди исследователей увеличился интерес к формализованному анализу мировых политических процессов и явлений, выросла методологическая вооруженность работ [Цветкова и др.

2023, с. 300]. Этот вывод подтверждают результаты исследования, проведенного учеными из Европейского университета в Санкт-Петербурге Ю.Г. Агафоновым и М.М. Соколовым [Агафонов, Соколов 2023]. Данные из опросов показали, что около половины респондентов за последние пять лет использовали статистические методы в исследовательской работе. При этом важно, что исследования, выполненные на основе количественной методологии, публикуются на страницах ведущих научных журналов [Вакарчук 2023; Дегтерев, Пискунов, Еремин 2023; Тимофеев 2022]. Помимо этого, успешно защищаются кандидатские и докторские диссертации, основанные на количественных методах анализа [Дегтерев 2019; Сытник 2022]. На площадках университетов, исследовательских институтов и экспертно-аналитических центров проводятся международные научные конференции и круглые столы методологического характера, посвященные опыту проведения исследований с опорой на математические методы обработки данных<sup>1</sup>. Более того, и это очень показательно, складывается впечатление, что отечественная международно-политическая наука созрела к началу методологического спора [Фененко 2018; Истомин, Байков, Худолей 2019; Дегтерев 2019; Фененко 2021] по аналогии со «вторым большим спором» в американской политической науке в 1950–1960-е гг., который развернулся между сторонниками формализованных методов анализа и защитниками традиционной историко-описательной методологии. По крайней мере очевидно, что среди ведущих отечественных исследователей-международников существуют диаметрально противоположные точки зрения относительно вопроса о том, в какую сторону должно двигаться методологическое развитие российской международно-политической науки.

Причины возросшего интереса отечественных исследователей к использованию количественных методов, безусловно, требуют отдельного подробного осмысления. Однако на первый взгляд представляется, что наиболее важным фактором является стреми-

---

<sup>1</sup> Семинар СМУиА по количественным методам // ИМЭМО РАН. 2022. URL: <https://www.imemo.ru/news/events/text/seminar-po-kolichestvennim-metodam> (дата обращения 3 декабря 2023); Международная конференция молодых ученых «Теория и методология науки о международных отношениях» // ИМЭМО РАН. 2022. URL: <https://www.imemo.ru/news/events/text/international-conference-for-young-researchers> (дата обращения 3 декабря 2023); Как преподавать международные отношения в условиях цифровизации: Лучшие практики РГГУ // РГГУ. 2021. URL: <https://www.rsuh.ru/education/fmopzr/news/detail.php?ID=719731> (дата обращения 3 декабря 2023).

тельное усложнение мировой политической системы, со свойственными этому процессу увеличением значимости и влияния негосударственных акторов, появлением новых форматов взаимодействия в новых областях, ускорением темпов социально-экономического развития, повсеместным внедрением информационно-коммуникационных технологий в сферы социального взаимодействия. В этой связи запрос современных исследователей-политологов на систематическое осмысление международно-политической среды с привлечением количественных методов обработки больших объемов информации представляется абсолютно логичным. Сопутствующим, но отнюдь не маловажным обстоятельством в этом отношении является цифровизация международно-политической реальности и международно-политической науки. Главное следствие этого процесса – рост числа доступных баз данных, в том числе текстовых, аккумулирующих в себе информацию о различных аспектах мировой политики.

В связи с этим существуют перечень публикаций отечественных специалистов-международников и политологов, в основе исследовательских программ которых лежит количественный анализ больших текстовых данных. Эти работы можно разделить на две группы. К первой группе относятся исследования, основанные на данных социальных сетей, преимущественно «Твиттер»\* (Социальная сеть «Твиттер» заблокирована на территории России по решению Генпрокуратуры от 24 февраля 2022 г. на основании ст. 15.3 закона об информации, информтехнологиях и о защите информации). Заслуживает внимания работа А.Н. Сытник [Цветкова, Сытник 2018], в которой автор на основе больших текстовых данных из социальной сети «Твиттер»\* изучает восприятие публичной дипломатии великих держав в зоне конфликтных столкновений. Особый акцент в исследовании сделан на использовании метода анализа тональности текстовой информации (*sentiment analysis*).

В коллективном исследовании А.С. Ахременко, Д.К. Стукала и А.П. Петрова [Ахременко, Стукал и Петров 2020] на основе big data из «Твиттер»\* определены факторы, идентифицирующие популярность твиттов\* в течение массовых политических протестов. Преобладание индикаторов, демонстрирующих сетевую позицию авторов твиттов\*, над индикаторами, определяющими содержание твиттов\*, идентифицировано авторами при помощи МНК-, и LASSO-регрессии, что является отличительным признаком этой работы.

Научные статьи из второй группы основаны на базах данных, аккумулирующих публикации мировых СМИ. В качестве таких баз чаще всего используются «Global database of events, languages

and tone”<sup>2</sup> (Далее – GDELT) и Factiva<sup>3</sup>. Следует оговориться, что существует принципиальная разница между этими базами. GDELT относится к типу баз, формируемых в автоматизированном режиме практически в реальном времени на основе кодирования публикаций мировых и локальных СМИ о политических, экономических, социальных процессах. Сбор и обработка новостной информации таким способом повышает надежность данных для исследований в области социальных наук. В базе закодированы данные о событиях с 1979 г. по настоящее время более чем на 100 языках, что делает GDELT одним из крупнейших существующих пространственно-временных наборов данных. Еще одним преимуществом этой базы данных является то, что все закодированные события содержат информацию о широте и долготе, что позволяет при необходимости идентифицировать местоположение каждого события [Bodas Sagi, Labeaga 2016, p. 38]. Помимо этого, база GDELT позволяет идентифицировать в рамках конкретных событий государства, людей, организации, темы, конкретные источники информации. Factiva, в свою очередь, представляет собой агрегатор, содержащий первичные новостные публикации мировых СМИ без их дополнительной предобработки. Кроме того, доступ к GDELT предоставляется в том числе на бесплатной основе, тогда как к Factiva – исключительно по платной подписке. Судя по всему, именно в силу этих причин большей популярностью у российских исследователей пользуется GDELT.

В первую очередь база GDELT востребована в отечественных публикациях для измерения динамики и качества межгосударственных отношений. Примером такого исследования является статья авторского коллектива под руководством Д.А. Дегтерева [Бадрутдинова, Дегтерев, Степанова 2017], в которой при помощи данных из GDELT была сделана попытка измерить отношения в треугольнике «США–КНР–РФ» в соответствии со шкалой Дж. Голдштейна. В результате анализа был идентифицирован ряд проблем. Во-первых, проблема полноты информации, содержащейся в GDELT. Авторы отмечают, что, несмотря на осуществление сбора новостей для GDELT более чем на ста языках, основной массив составляют англоязычные публикации, что может исказить результаты исследований. Во-вторых, акцентируется внимание на проблеме непопадания в ленты новостных агентств данных о внеш-

---

<sup>2</sup> The GDELT Project // 2023. URL: <https://www.gdeltproject.org> (дата обращения 3 декабря 2023).

<sup>3</sup> Factiva // Dow Jones. 2023. URL: <https://www.dowjones.com/professional/factiva/> (дата обращения 3 декабря 2023).

неторговых, инвестиционных и миграционных потоках. Таким образом, авторы приходят к выводу, что для получения более точных результатов измерения отношений государств за определенный период необходимо использовать альтернативную методологию, основанную на открытых статистических данных [Бадрутдинова, Дегтерев и Степанова 2017, с. 85]. Важно отметить при этом, что тезис авторов об отсутствии полноты данных GDELT не может носить обобщающий характер и зависит от контекста исследования. Кроме того, интересно то, что авторская гипотеза была верифицирована как при использовании ивент-анализа на основе данных GDELT, так и при применении альтернативной методологии.

Еще одна проблема работы с GDELT обусловлена лежащей в ее основе онтологией CAMEO (conflict and mediation event observations)<sup>4</sup>. Эта система представляет собой иерархическую классификацию событий, в которой каждое типовое событие имеет свой код и сопровождающее описание. Это означает, что каждая публикация СМИ перед занесением в базу автоматически кодируется и распределяется к той или иной укрупненной категории событий<sup>5</sup>. Онтология CAMEO предусматривает 300 таких категорий. Помимо событий, содержащихся в новостном заголовке, алгоритмы GDELT также автоматически кодируют акторов и географические локации. Для понимания принципа кодирования новостной информации рассмотрим пример новостного заголовка, проанализированного алгоритмом GDELT: “Turkey will allow up to 13,000 Turkish Kurd refugees who have lived in Iraq for more than a decade to return home as part of a UN-brokered deal”<sup>6</sup>. Красным цветом в этом примере выделен действующий актор, синим – актор-реципиент, зеленым – событие.

Несмотря на то что создание онтологий позволило осуществлять автоматический сбор big data о международных событиях для их последующего количественного и качественного анализа, любая из существующих в настоящий момент онтологий априори несовершенна. Это приводит к тому, что зачастую отличные друг от друга по своей сути события могут группироваться в процессе автоматического кодирования в одну категорию [Schrodt, Van Brackle 2013, pp. 23–49]. Именно поэтому в настоящий момент разрабатывается новая система иерархической классификации

---

<sup>4</sup> *Schrodt P. CAMEO Event and actor Codebook // The GDELT Project. 2023. URL: <http://data.gdelproject.org/documentation/CAMEO.Manual.1.1b3.pdf> (дата обращения 3 декабря 2023).*

<sup>5</sup> Ibid.

<sup>6</sup> Ibid.

событий PLOVER (political language ontology for verifiable event records), которая должна заменить CAMEO [Halterman et al. 2023].

В исследовании автора настоящей статьи, опубликованном в 2021 г. [Вакарчук 2021], был использован один из базовых способов преодоления проблемы несовершенства онтологии CAMEO. В этой работе наряду с применением машинного ивент-анализа новостные публикации о российско-польском взаимодействии по проблеме исторической памяти о событиях Второй мировой войны отбирались, кодировались и взвешивались по шкале Дж. Голдштейна вручную по причине отсутствия соответствующей категории в CAMEO. Разумеется, такой вариант не всегда является эффективным и скорее возможен в узкопрофильных исследованиях, которые не требуют обращения к big data. Однако, несмотря на обозначенную проблему валидности данных GDELT, возможность автоматизации сбора данных позволяет более эффективно работать с действительно репрезентативными выборками.

Данные GDELT также имеют большой потенциал для проведения на их основе более сложных исследований с использованием методов математической статистики. Примечательными в этом отношении для отечественной научной литературы являются исследования в русле политической компаративистики. Примером такого исследования является другая наша работа [Вакарчук 2020], в которой количественные методы анализа текста сочетаются с методами математической статистики.

Для оценки влияния различий между государствами-членами ЕС на качество их отношений с РФ автор операционализировал зависимую переменную – качество отношений между государствами-членами ЕС и РФ – посредством данных GDELT. Таким образом, метод ивент-анализа был применен автором для решения лишь одной из исследовательских задач. В качестве предикторов автор использовал ряд переменных, идентифицирующих неоднородную структуру ЕС. Для решения задачи выявления взаимосвязи между переменными был применен статистический U-критерий Манна–Уитни. Выбранный исследовательский дизайн продемонстрировал наличие статистически значимого влияния предиктора принадлежности/отсутствия принадлежности государств-членов ЕС к великим державам на их отношения с РФ [Вакарчук 2020, с. 69].

В работе М.А. Завадской [Завадская 2015] данные GDELT были использованы для сопоставительного анализа протестной активности и частоты репрессий в период электоральных циклов в зависимости от типа институционального дизайна в 64 авторитарных государствах. Последующая обработка данных при помощи методов описательной статистики наглядно продемонстрировала

разницу между персоналистскими и партийными авторитарными режимами. В исследовании был представлен вывод, что персоналистские режимы со слабыми институтами более склонны к использованию репрессий в качестве политического инструмента и, следовательно, к утрате легитимности и массовым протестам в электоральные циклы [Завадская 2015, с. 21].

Иное применение данных GDELТ представлено в работе Н.В. Ярцевой [Ярцева 2021]. Это исследование имеет скорее практико-ориентированный характер. Исследователь поставила цель на конкретных примерах продемонстрировать несколько способов обращения с большими данными GDELТ: при помощи языков Python и SQL, а также инструментария платформы GDELТ Summary<sup>7</sup>. SQL был использован автором для извлечения выборки из Google Cloud Public Datasets, сервиса, где хранятся данные GDELТ. При помощи языка Python полученные автором выборки были обработаны, визуализированы и проверены на предмет статистической значимости. GDELТ Summary является альтернативной для обращения к данным GDELТ и не требует написания кода. Для демонстрации возможностей GDELТ Summary автором был использован один из инструментов этого сервиса – Television Explorer<sup>8</sup>. Этот инструмент позволяет формировать запросы к базе телевизионных новостей GDELТ и визуализировать результаты запроса.

Цель автора заключалась в том, чтобы выявить более негативное отношение к движению «желтых жилетов» во французской прессе, чем в итальянской. Для этого автор извлекла из GDELТ выборку публикаций мировых СМИ, освещавших протесты в Париже в 2018–2019 гг. SQL запрос был сформирован так, чтобы отфильтровать базу данных по коду CAMEO 175 “Use repression” и территории событий соответственно с группировкой по дате. Для доказательства гипотезы автор использовала статистический t-критерий Стьюдента.

Несмотря на интерес российских политологов к прикладным исследованиям и методологическим проблемам использования данных GDELТ, практика последних лет складывается так, что зарубежные исследования по аналогичной тематике более обширны и основательны. Зарубежные публикации отличает использование авторами широкого круга методов математической статистики для

---

<sup>7</sup> GDELТ summary. URL: <https://blog.gdelproject.org/announcing-gdel-summary/> (дата обращения 8 сентября 2023).

<sup>8</sup> Television Explorer. URL: <https://api.gdelproject.org/api/v2/summary/summary?d=iatv&t=summary> (дата обращения 8 сентября 2023).

изучения социальных процессов в политической плоскости. Эту тенденцию можно объяснить ограниченностью описательных методов, используемых для анализа больших объемов информации, полученных машинным путем.

Значимое место среди этих работ занимают исследования, ориентированные на проверку гипотез о наличии связей между переменными. Публикация М. Куаранты [Quaranta 2016] – пример работы, направленной на изучение протестной активности при помощи количественной методологии и big data. Автор сделал попытку оценить влияние экономического кризиса 2008–2014 гг. на устойчивость десяти европейских демократий. Эффект влияния экономического кризиса был измерен при помощи линейной МНК-регрессии. Это позволило проанализировать влияние изменения экономической ситуации на верховенство закона, чувствительность граждан к правительственным инициативам и на стабильность политической системы. Одна из зависимых переменных указанной модели «неконвенциональное политическое участие граждан» была операционализирована при помощи GDELT. Для этого были собраны данные о частоте освещения мировыми СМИ протестной активности в европейских странах в указанный период.

Помимо МНК-регрессии для моделирования зависимых переменных, принимающих только натуральные значения, зарубежными авторами часто используется метод отрицательной биномиальной регрессии. Примером такого исследования является работа Й. Бунте и Л. Винсон [Bunte, Vinson 2016], в которой авторы при помощи указанного метода определяют влияние неформальных соглашений о разделении властей на местном уровне на степень межрелигиозного насилия между мусульманами и христианами в Нигерии. Применение этого метода продемонстрировало, что уровень межрелигиозного насилия значительно ниже в районах с функционирующими институтами разделения властей, чем в районах без таких институтов [Bunte, Vinson 2016, pp. 63–64].

В этой работе данные GDELT использовались для операционализации конфликтности между мусульманами и христианами на территории Нигерии за период с 2000 по 2006 г. Примечательно, что в исследовании отмечается проблема качества содержащихся в GDELT данных применительно к верификации выдвинутой гипотезы. Автор указывает, что выборка новостных публикаций, извлеченная из GDELT, не в полной мере соотносится с собранными и кодированными вручную новостными публикациями нигерийских локальных газет [Bunte, Vinson 2016, p. 52]. Таким образом, вновь в литературе подчеркивается необходимость осто-

рожного обращения с эмпирическими данными GDELT в прикладных исследованиях. Между тем результаты проведенного автором регрессионного анализа продемонстрировали свою валидность как при использовании в качестве зависимой переменной событийных данных, собранных вручную, так и при использовании результатов машинного ивент-анализа.

Помимо проблемы полноты данных, в зарубежных публикациях акцентируется внимание на технических ограничениях работы с GDELT. В исследовании К. Джазери [Jazayeri 2016], в котором оценивается влияние политического неравенства, основанного на идентичности, на вероятность ненасильственного/насиленного сопротивления на Ближнем Востоке и в Северной Африке, представлено обоснование отказа от использования GDELT. Автор мотивировал это решение тем, что платформа GDELT Analytical Service<sup>9</sup>, предлагающая набор инструментов для работы с данными GDELT без написания кода, не функционировала на момент проведения исследования [Jazayeri 2016, p. 19]<sup>10</sup>.

Иное применение данных GDELT представлено в относительно недавней работе исследовательского коллектива под руководством К. Дэвис о влиянии политической сферы межгосударственных отношений на международную торговлю на примере отношений Китая и Индии с их ведущими торговыми партнерами [Davis, Fuchs, Johnson 2019]. Для верификации гипотезы авторы измерили уровень отношений в политической сфере между указанными странами в период с 1990 по 2012 г. при помощи данных GDELT, исключив из итоговой выборки положительно взвешенные события по шкале Дж. Голдштейна. Более того, для идентификации различий влияния этих акторов на торговые отношения выборка была ограничена также подсчетом негативных действий, инициатором которых выступали военные и государственные акторы торговых партнеров. Валидность операционализации была установлена посредством оценки совпадения позиций Китая и Индии с государствами-партнерами в Генеральной Ассамблее ООН. В свою очередь, зависимая переменная была разбита на два типа с учетом информации о государственном контроле за внешнеторговым оборотом.

В результате использования гравитационной модели внешней торговли авторы пришли к выводу, что импорт, контролируемый государственными предприятиями, более чувствителен к изме-

---

<sup>9</sup> GDELT Analysis Service. URL: <https://analysis.gdeltproject.org> (дата обращения 8 сентября 2023).

<sup>10</sup> На момент написания текста доклада сервис по-прежнему не работает.

нению климата двусторонних отношений в политической сфере, чем импорт, контролируемый частными предприятиями. Такая закономерность была характерна как для Китая, так и для Индии. Аналогичная закономерность была обнаружена для индийского, но не для китайского экспорта. Авторы объясняют это большей независимостью китайских экспортно-ориентированных предприятий от внешнеполитического курса государства в исследуемый период [Davis, Fuchs, Johnson 2019, pp. 429–430].

На основе этих выводов было представлено предположение о росте политизации внешней торговли по мере увеличения количества стран с частично или полностью контролируемой государством экономикой. Это предположение примечательно относительно продолжающейся дискуссии о роли государства в мировой политике в контексте глобализации.

Данные GDELT могут быть использованы также для оценки внутри- и внешнеполитических рисков. Интересный подход в этом отношении был предложен китайскими исследователями-международниками [Sun, Gao, Liu, Wang 2021]. Авторы не только содержательно оценили возможности и ограничения использования данных GDELT для идентификации страновых политических рисков в период с 2013 по 2019 г., но и выявили взаимосвязь между политическими рисками и объемом китайских инфраструктурных инвестиций на пространстве проекта «Один пояс – один путь».

Новизна авторского подхода заключалась в предложенной формуле расчета индекса политического риска через отношение модуля суммы событий класса GDELT «Материальный конфликт»<sup>11</sup>, взвешенных по шкале Дж. Голдштейна, к сумме всех событий, взвешенных по шкале Дж. Голдштейна за определенный период. На основе набора данных в работе продемонстрировано, что указанный алгоритм способен достовернее показывать взаимосвязь политического риска и объема китайских инвестиций, чем простое использование количества событий той или иной укрупненной категории CAMEO. При этом доказана неэффективность использования событий укрупненной категории «Протест» для цели исследования, так как выбранные авторами 63 государства-участника китайского проекта отличаются друг от друга по типу политического режима.

---

<sup>11</sup> Согласно онтологии CAMEO, класс «Материальный конфликт» включает в себя 6 укрупненных категорий: «Протест», «Демонстрация силы», «Принуждение», «Нападение», «Бои», «Применение неконвенционального массового насилия».

В результате исследователи пришли к выводу, что политические риски положительно коррелируют с объемом китайских инфраструктурных инвестиций, что противоречит практике увеличения объема инвестиций по мере снижения политических рисков. При этом для пространства проекта «Один пояс – один путь» в большей степени были характерны внешнеполитические, а не внутривнутриполитические риски в рамках исследуемого периода [Sun, Gao, Liu, Wang 2021].

### *Заключение*

Сопоставляя практику использования данных GDELT в отечественных и зарубежных исследованиях, можно прийти к следующим выводам. В научных работах отечественных исследователей данные из этой базы используются преимущественно в описательных целях, без дополнительной предобработки. Для работ зарубежных ученых характерна инкорпорация данных GDELT в сложные математические модели с целью определения характера и направления связей между переменными.

Проведенный обзор позволяет говорить о разнообразии исследовательских задач, которые позволяют решить данные GDELT: оценка объема и качества освещения СМИ той или иной проблемы внешне- и внутривнутриполитического характера, измерение динамики и качества межгосударственных или межконфессиональных отношений, частоты протестной активности в той или иной стране. Такая исследовательская вариативность обусловлена широким функционалом, которым обладает база GDELT.

К ключевым преимуществам использования базы данных GDELT для решения исследовательских задач можно отнести, во-первых, возможность проведения автоматизированного ивент- и контент-анализа больших текстовых данных, во-вторых, дополнительную опцию для операционализации и валидации политических явлений и процессов, в-третьих, возможность для фальсификации результатов проведенного анализа. Между тем, как в отечественной, так и в зарубежной литературе, зафиксирован ряд ограничений работы с данными GDELT: во-первых, несовершенство онтологии SAMEO, во-вторых, неполнота данных, содержащихся в базе, в-третьих, технические ограничения работы с базой GDELT. Соответственно применять напрямую данные GDELT следует крайне осторожно, тестируя в каждом конкретном случае эти измерения на валидность.

## Литература

---

- Агафонов, Соколов 2023 – *Агафонов Ю.Г., Соколов М.М.* Российская политология в 2021 г.: социальный и интеллектуальный ландшафт // Полис: Политические исследования. 2023. № 2. С. 54–71.
- Ахременко, Стукал, Петров 2020 – *Ахременко А.С., Стукал Д.К., Петров А.П.* Сеть или текст? Факторы распространения протеста в социальных медиа: теория и анализ данных // Полис: Политические исследования. 2020. № 2. С. 73–91.
- Бадрутдинова, Дегтерев, Степанова 2017 – *Бадрутдинова К.Р., Дегтерев Д.А., Степанова А.А.* Отношения в треугольнике США – РФ – КНР: соблюдается ли формула лидерства Г. Киссинджера? // Вестник международных организаций. 2017. Т. 12. № 1. С. 81–109.
- Вакарчук 2020 – *Вакарчук Д.О.* Влияние различий между государствами-членами ЕС на их отношения с Российской Федерацией: 2014–2019 гг. // Вестник РГГУ. Серия «Политология. История. Международные отношения». 2020. № 4. С. 57–73.
- Вакарчук 2021 – *Вакарчук Д.О.* Вторая мировая война в контексте современных российско-польских отношений // Вестник РГГУ. Серия «Политология. История. Международные отношения». 2021. № 2. С. 80–90.
- Вакарчук 2023 – *Вакарчук Д.О.* Эмпирическая верификация концепции регионального лидерства на примере восходящих держав // Вестник Санкт-Петербургского университета. Международные отношения. 2023. Т. 16. № 3. С. 255–274.
- Дегтерев 2015 – *Дегтерев Д.А.* Количественные методы в международных исследованиях // Международные процессы. 2015. № 2 (41). С. 35–54.
- Дегтерев 2019 – *Дегтерев Д.А.* «Второй большой спор» в контексте становления российской науки о международных отношениях // Международные процессы. 2019. Т. 17. № 2 (57). С. 43–63.
- Дегтерев 2019 – *Дегтерев Д.А.* Количественные методы в прикладном анализе международных отношений: Дис. ... д-ра полит. наук. М.: Гос. ун-т им. М.В. Ломоносова, 2019. 644 с.
- Дегтерев, Пискунов, Еремин 2023 – *Дегтерев Д.А., Пискунов Д.А., Еремин А.А.* 5G-конкуренция США и КНР в странах Латинской Америки: у истоков технологического декаплинга // Полис: Политические исследования. 2023. № 3. С. 20–38.
- Истомин, Байков, Худолей 2019 – *Истомин И.А., Байков А.А., Худолей К.К.* Международные отношения: наука без метода // Международные процессы. 2019. Т. 17. № 2 (57). С. 63–90.
- Завадская 2015 – *Завадская М.А.* Время выбирать: манипулирование электоральными циклами в современных авторитарных режимах // Вестник Пермского университета. Серия «Политология». 2015. № 4. С. 5–28.
- Сытник 2022 – *Сытник А.Н.* Цифровизация и «большие данные» в международных отношениях: теоретические, методологические и прикладные аспекты: Дис. ... канд. полит. наук. СПб.: Санкт-Петербургский гос. ун-т, 2022. 479 с.

- Тимофеев 2022 – *Тимофеев И.Н.* Политика санкций США на уровне исполнительной власти // *Мировая экономика и международные отношения*. 2022. Т. 66. № 3. С. 23–32.
- Фененко 2021 – *Фененко А.В.* Иллюзорная дисциплина: и снова о количественных методах в международных отношениях // *Международные процессы*. 2021. Т. 19. № 4 (67). С. 163–191.
- Фененко 2018 – *Фененко А.В.* Статистика против истории // *Международные процессы*. 2018. Т. 16. № 3 (54). С. 56–83.
- Цветкова, Сытник 2018 – *Цветкова Н.А., Сытник А.Н.* Публичная дипломатия в Афганистане, 2002–2018 гг.: влияние США на социальные сети, политическую систему и университеты // *Вестник Санкт-Петербургского университета. Международные отношения*. 2018. Т. 11. Вып. 4. С. 344–361.
- Цветкова и др. 2023 – *Цветкова Н.А., Худoley К.К., Хейфец В.Л. и др.* Научные журналы в условиях трансформации международных отношений: проблемы, вызовы и перспективы. Часть 1 / Н.А. Цветкова, К.К. Худoley, В.Л. Хейфец, В.Н. Гарбузов, И.А. Истомин, М.В. Харкевич // *Вестник Санкт-Петербургского университета. Международные отношения*. Т. 16. № 3. С. 294–315.
- Ярцева 2021 – *Ярцева Н.В.* Анализ больших объемов данных: возможности Gdelt Project при использовании языка программирования Python: Опыт гуманитария, решившего постичь Big Data // *Мониторинг общественного мнения: экономические и социальные перемены*. 2021. № 1. С. 351–367.
- Bodas Sagi, Labeaga 2016 – *Bodas Sagi D., Labeaga J.* Using GDELТ data to evaluate the confidence on the Spanish government energy policy // *International Journal of Interactive Multimedia and Artificial Inteligence*. 2016. No. 3 (6). P. 38–43.
- Bunte, Vinson 2016 – *Bunte J., Vinson L.* Local power-sharing institutions and interreligious violence in Nigeria // *Journal of Peace Research*. 2016. No. 1 (53). P. 49–65.
- Davis, Fuchs, Johnson 2019 – *Davis C., Fuchs A., Johnson K.* State control and the effects of foreign relations on bilateral trade // *Journal of Conflict Resolution*. 2019. No. 2 (63). P. 405–438.
- Halterman et al. 2023 – *Halterman A., Bagozzi B., Beger B., Schrodт P., Scarborough G.* PLOVER and POLECAT. A new political event ontology and dataset // SocArXiv. 2023. URL: <https://osf.io/preprints/socarxiv/rm5dw/download> (дата обращения 25 марта 2024). DOI: 10.31235/osf.io/rm5dw.
- Jazayeri 2016 – *Jazayeri K.* Identity-based political inequality and protest. The dynamic relationship between political power and protest in the Middle East and North Africa // *Conflict Management and Peace Science*. 2016. No. 4 (33). P. 400–422.
- Quaranta 2016 – *Quaranta M.* Protesting in “hard times”. Evidence from a comparative analysis of Europe, 2000–2014 // *Current Sociology*. 2016. No. 5 (64). P. 736–756.
- Schrodт, Van Brackle 2013 – *Schrodт P., Van Brackle D.* Automated coding of political event data // *Handbook of computational approaches to counterterrorism* / Ed. by V. Subrahmanian. N.Y.: Springer, 2013. P. 23–49.
- Sun, Gao, Liu, Wang 2021 – *Sun X., Gao J., Liu B., Wang Z.* Big data-based assessment of political risk along the belt and road // *Sustainability*. 2021. № 13 (7). URL: <https://doi.org/10.3390/su13073935>.

## References

---

- Agafonov, Yu.G. and Sokolov, M.M. (2023), "Russian political science in 2021. Social and intellectual landscape", *Polis. Political Studies*, no. 2, pp. 54–71.
- Ahremenko, A.S., Stukal, D.K. and Petrov, A.P. (2020), "Web or text? Factors of protest spreading in social media. Theory and data analysis", *Polis. Political Studies*, no. 2, pp. 73–91.
- Badrutdinova, K.R., Degterev, D.A. and Stepanova, A.A. (2017), "Interconnections among the United States, Russia and China. Whether G. Kissinger's leadership formula is being followed?", *International Organisations Research Journal*, vol. 12, no. 1, pp. 81–109.
- Bodas-Sagi, D. and Labeaga, J. (2016), "Using GDELT data to evaluate the confidence on the Spanish government energy policy", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 3, pp. 38–43.
- Bunte, J. and Vinson, L. (2016), "Local power-sharing institutions and interreligious violence in Nigeria", *Journal of Peace Research*, vol. 53, no. 1, pp. 49–65.
- Davis, C., Fuchs, A. and Johnson, K. (2019), "State control and the effects of foreign relations on bilateral trade", *Journal of Conflict Resolution*, vol. 63, no. 2, pp. 405–438.
- Degterev, D.A. (2015), "Quantitative methods in international studies", *International Trends*, vol. 41, no. 2, pp. 35–54.
- Degterev, D.A. (2019), *Kolichestvennye metody v prikladnom analize mezhdunarodnykh otnoshenii* [Quantitative methods in the applied analysis of international relations], D. Sc. Thesis (Political Science), Moscow State University, Moscow, Russia.
- Degterev, D.A. (2019), "‘The Second Great Controversy’ in the context of the formation of the Russian science on international relations", *International Trends*, vol. 17, no. 2 (57), pp. 43–63.
- Degterev, D.A., Piskunov, D.A. and Eremin, A.A. (2023), "U.S. – China 5G-rivalry in Latin America. At the origins of technological decoupling", *Polis. Political Studies*, no. 3, pp. 20–38.
- Fenenko, A.V. (2021), "The illusory discipline. Once again on quantitative methods in international relations", *International Trends*, vol. 19, no. 4 (67), pp. 163–191.
- Fenenko A.V. (2018), "Statistics versus History", *International*, vol. 16, no. 3 (54), pp. 56–83.
- Halterman, A., Bagozzi, V., Beger, V., Schrodtt, P.A. and Scraborough, G. "PLOVER and POLECAT. A new political event ontology and dataset", in SocArXiv. 2023, available at: <https://osf.io/preprints/socarxiv/rm5dw/download> (Accessed 25 March 2024), DOI: 10.31235/osf.io/rm5dw.
- Istomin, I.A., Baikov, A.A. and Khudolei, K.K. (2019), "International relations. Science without method?", *International Trends*, vol. 17, no. 2 (57), pp. 63–90.
- Jazayeri, K. (2016), "Identity-based political inequality and protest. The dynamic relationship between political power and protest in the Middle East and North Africa", *Conflict Management and Peace Science*, vol. 33, no. 4, pp. 400–422.

- Quaranta, M. (2016), "Protesting in 'hard times'. Evidence from a comparative analysis of Europe, 2000–2014", *Current Sociology*, vol. 64, no. 5, pp. 736–756.
- Schrodt, P.A. and Van Brackle, D. (2013), "Automated coding of political event data", in Subrahmanian, V., ed., *Handbook of computational approaches to counterterrorism*, Springer, New York, USA, pp. 23–49.
- Sun, X., Gao, J., Liu, B. and Wang, Z. (2021), "Big data-based assessment of political risk along the belt and road", *Sustainability*, vol. 13, no. 7. URL: <https://doi.org/10.3390/su13073935>.
- Sytnik, A.N. (2022), Tsifrovizatsiya i "bol'shie dannye" v mezhdunarodnykh otnosheniyakh: teoreticheskie, metodologicheskie i prikladnye aspekty [Digitalization and "big data" in international relations. Theoretical, methodological and applied aspects], Ph.D. Thesis (Political Science), Sintt Petersburg University, Saint Petersburg, Russia.
- Timofeev, I.N. (2022), "An executive power dimension of the U.S. Sanctions policy", *Mirovaya ekonomika i mezhdunarodnye otnosheniya*, vol. 66, no. 3, pp. 23–32.
- Tsvetkova, N.A. and Sytnik, A.N. (2018), "Public diplomacy in Afghanistan, 2002–2018. Impact of the USA on social networks, political system and universities", *Vestnik Sankt-Peterburgskogo universiteta. Mezhdunarodnye otnosheniya*, vol. 11, iss. 4, pp. 344–361.
- Tsvetkova, N.A., Khudoley, K.K., Jeifets, V.L., Garbuzov, V.N., Istomin, I.A. and Kharkevich, M.V. (2023), "Scientific journals in the context of the transformation of international relations. Issues, challenges and prospects. Part 1", *Vestnik Sankt-Peterburgskogo universiteta. Mezhdunarodnye otnosheniya*, vol. 16, no. 3, pp. 294–315.
- Vakarchuk, D.O. (2023), "Empirical verification of the regional leadership concept on the example of rising powers", *Vestnik Sankt-Peterburgskogo universiteta. Mezhdunarodnye otnosheniya*, vol. 16, no. 3, pp. 255–274.
- Vakarchuk, D.O. (2020), "Differences among the member states of the European Union and its influence on their relations with the Russian Federation: 2014–2019", *RSUH/RGGU Bulletin. "Political Science. History. International Relations" Series*, no. 4, pp. 57–73.
- Vakarchuk, D.O. (2021), "World war II in the context of contemporary Russian-Polish relations", *RSUH/RGGU Bulletin. "Political Science. History. International Relations" Series*, no. 2, pp. 80–90.
- Yartseva, N.V. (2021), "Analysing large amounts of data. GDELT project's opportunities using the Python programming language. A humanities scholar's experience with Big Data, *Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny*, no. 1, pp. 351–367.
- Zavadsкая, M.A. (2015), "Time to choose. Manipulating electoral cycles in present-day authoritarian regimes", *Vestnik Permskogo universiteta. Seriya: Politologiya*, no. 4, pp. 5–28.

*Информация об авторе*

*Денис О. Вакарчук*, кандидат исторических наук, Российский государственный гуманитарный университет, Москва, Россия; 125047, Россия, Москва, Миусская пл., д. 6; vakarchuk.d@rggu.ru

*Information about the author*

*Denis O. Vakarchuk*, Cand. of Sci. (History), Russian State University for the Humanities, Moscow, Russia; 6, Miusskaya Sq., Moscow, Russia, 125047; vakarchuk.d@rggu.ru