

МЕТОДЫ И МЕТОДОЛОГИЯ

DOI: 10.14515/monitoring.2019.2.03

Правильная ссылка на статью:

Яковлева Д. А., Колесниченко О. Ю., Мазелис Л. С., Григоревский И. Н., Колесниченко Ю. Ю. API-социология и глобальное сообщество Google: живая система социальной трансформации // Мониторинг общественного мнения: Экономические и социальные перемены. 2019. № 2. С. 54—79. <https://doi.org/10.14515/monitoring.2019.2.03>.

For citation:

Yakovleva D. A., Kolesnichenko O. Yu., Mazelis L. S., Grigorevsky I. N., Kolesnichenko Y. Yu. (2019) Api-sociology and Google global community: the live system of social transformation. *Monitoring of Public Opinion: Economic and Social Changes*. No. 2. P. 54—79. <https://doi.org/10.14515/monitoring.2019.2.03>.



**Д. А. Яковлева, О. Ю. Колесниченко, Л. С. Мазелис,
И. Н. Григоревский, Ю. Ю. Колесниченко**

API-СОЦИОЛОГИЯ И ГЛОБАЛЬНОЕ СООБЩЕСТВО GOOGLE: ЖИВАЯ СИСТЕМА СОЦИАЛЬНОЙ ТРАНСФОРМАЦИИ

API-СОЦИОЛОГИЯ И ГЛОБАЛЬНОЕ СООБЩЕСТВО GOOGLE: ЖИВАЯ СИСТЕМА СОЦИАЛЬНОЙ ТРАНСФОРМАЦИИ

API-SOCIOLOGY AND GOOGLE GLOBAL COMMUNITY: THE LIVE SYSTEM OF SOCIAL TRANSFORMATION

ЯКОВЛЕВА Дарья Алексеевна — аспирант, Владивостокский государственный университет экономики и сервиса, Владивосток, Россия
E-MAIL: darya.yakovleva15@vvsu.ru
<https://orcid.org/0000-0002-0139-4051>

Dariya A. YAKOVLEVA¹ — Post-Graduate Student
E-MAIL: darya.yakovleva15@vvsu.ru
<https://orcid.org/0000-0002-0139-4051>

¹ Vladivostok State University of Economics and Service, Vladivostok, Russia

КОЛЕСНИЧЕНКО Ольга Юрьевна — кандидат медицинских наук, старший преподаватель, Первый Московский государственный медицинский университет имени И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия
E-MAIL: oykolesnichenko@list.ru
<https://orcid.org/0000-0002-4523-6485>

МАЗЕЛИС Лев Соломонович — доктор экономических наук, профессор, заведующий кафедрой, Владивостокский государственный университет экономики и сервиса, Владивосток, Россия
E-MAIL: lev.mazelis@vvsu.ru
<https://orcid.org/0000-0001-7346-3960>

ГРИГОРЕВСКИЙ Иван Николаевич — кандидат технических наук, научный сотрудник, Институт программных систем имени А.К. Айламазяна Российской академии наук, Переславль-Залесский, Россия; ответственный секретарь оргкомитета, Национальный Суперкомпьютерный Форум, Переславль-Залесский, Россия
E-MAIL: gin@nscf.ru
<https://orcid.org/0000-0001-8078-1238>

КОЛЕСНИЧЕНКО Юрий Юрьевич — директор, Интернет-портал «Бюллетень «Анализ безопасности», Москва, Россия
E-MAIL: green-apple_2000@mtu-net.ru
<https://orcid.org/0000-0002-4426-9250>

Olga Yu. KOLESNICHENKO² — *Cand. Sci. (Med.), Senior Lecturer*
E-MAIL: oykolesnichenko@list.ru
<https://orcid.org/0000-0002-4523-6485>

Lev S. MAZELIS¹ — *Dr. Sci. (Econ.), Professor, Head of Department*
E-MAIL: lev.mazelis@vvsu.ru
<https://orcid.org/0000-0001-7346-3960>

Ivan N. GRIGOREVSKY^{3,4} — *Cand. Sci. (Tech.), Research Fellow; Executive Secretary of the Organizing Committee*
E-MAIL: gin@nscf.ru
<https://orcid.org/0000-0001-8078-1238>

Yuriy Yu. KOLESNICHENKO⁵ — *Chief*
E-MAIL: green-apple_2000@mtu-net.ru
<https://orcid.org/0000-0002-4426-9250>

¹ Vladivostok State University of Economics and Service, Vladivostok, Russia

² I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, Russia

³ Aylamazyan Program Systems Institute of Russian Academy of Sciences, Pereslavl-Zalesky, Russia

⁴ National Supercomputing Forum, Pereslavl-Zalesky, Russia

⁵ Security Analysis Bulletin, Moscow, Russia

Аннотация. В данной статье авторы представляют результаты морфологического текстового анализа больших

Abstract. In this paper, the authors present the results of morphological text big data analysis for the occurrence

данных на предмет встречаемости в интернете ключевых слов для различных стран. Значения для слов находились при запросе из баз данных поисковой системы Google через API-доступ при помощи специального программного обеспечения, созданного во Владивостокском государственном университете экономики и сервиса. Проведена кластеризация стран, а также ключевых слов методом k-means и иерархическим методом. Морфологический текстовый анализ позволяет оценить глобальный образ мировой интернет-дискуссии, а для каждой из стран — степень адаптации к условиям нового технологического уклада и уровень социального стресса. Построена модель глобальной дискуссии социальной направленности, что соотносится с трендами социальной нестабильности по данным ежегодного доклада «The Global Risks Report» Всемирного экономического форума. Проведен тематический анализ трех потенциалов развития по Р.С. Гринбергу (природному, интеллектуальному, пространственному), который показал, что, по сравнению с рядом стран, Россия занимает устойчивые и перспективные позиции.

Ключевые слова: текстовая аналитика, Большие данные, визуализация, кластерный анализ, API-социология, наркомания

of keywords over the Internet for different countries. The values for the words were requested from the databases of the Google search engine via API-access using special software created at the Vladivostok State University of Economics and Service. The countries were analyzed using cluster analysis, the keywords - using k-means and hierarchical method. Using morphological text analytics, it is possible to evaluate the image of the global Internet discussion and to estimate the level of adaptation to the conditions of the new technological order and level of social stress for each country. A social Global Discussion Model was constructed, which corresponds to the social instability trends according to “The Global Risks Report” provided by the World Economic Forum. Authors completed morphological Big Data text analysis of the three types of development potential by R. S. Greenberg (natural, intellectual, spatial), which showed that Russia occupies a stable and promising position in comparison with a number of countries.

Keywords: text analytics, Big Data, Google, visualization, Data Mining, Cluster analysis, API-sociology, drug addiction

Введение

В статье представлены результаты морфологической текстовой аналитики Больших данных — выбранных для анализа слов, доступных при запросе из баз данных поисковой системы Google. Для осуществления такого запроса к Google и подсчета количества заданных (ключевых) слов через API-доступ (Application Programming Interface) во Владивостокском государственном университете экономики и сервиса была разработана специальная компьютерная программа. Это уже третий этап аналитики, выполненный с помощью созданного программного

обеспечения, результаты двух предыдущих этапов были опубликованы в журнале «Мониторинг общественного мнения: Экономические и социальные перемены» [Колесниченко и др., 2015; Колесниченко и др., 2016]. Данный вид аналитики относится к области API-социологии и опирается на открытые ресурсы интернета. Исследование выполняется под эгидой Национального Суперкомпьютерного Форума — Национальной Суперкомпьютерной Технологической Платформы. Результаты докладывались на ежегодном форуме, аналитические подходы корректировались после обсуждения докладов с научной аудиторией и учета замечаний. В 2018 г. результаты данного исследования по API-социологии были представлены в докладе на VIII Грушинской конференции ВЦИОМ, что стало важным этапом в становлении этого нового научного социологического направления.

Интернет рассматривается как глобальная система «People-to-IT» (люди вкладывают в информационные системы черты своего бытия), обладающая гибкостью с точки зрения реакции на события, происходящие в социуме. Повторим кратко описание основных подходов к пониманию самой методики, изложенное в ранее опубликованных статьях. Интернет является площадкой для дискуссии глобальной аудитории. Искусственно исказить на глобальном уровне информационные следы этой дискуссии невозможно, по крайней мере, на данном этапе технологического развития. О полностью контролируемом информационном пространстве можно будет говорить только если возникнет глобальная система «IT-to-People», когда алгоритмы искусственного интеллекта на упреждающей скорости будут способны управлять социальной системой людей с контролем любого информационного следа от человека в информационных сетях.

Количество (частота встречаемости) тех или иных ключевых слов в текстах интернета отражает разные аспекты в жизни людей: психологические, стрессовые, социальные, рекламные, рыночные, политические. Поисковая система Google фиксирует всевозможные тексты: новостные, рекламные, аналитические, блоги, комментарии. Мы исходим из того, что через частоту встречаемости ключевых слов можно оценивать те или иные процессы, особенно если слова подсчитываются в привязке к названиям стран и определенным годам.

В привязке к названиям стран на количество ключевых слов влияет множество факторов: число компьютеров на душу населения страны; уровень компьютерной грамотности населения; число IP-соединений и доступность интернета в стране в зависимости от политических или экономических условий; общий уровень экономического развития в стране и число интернет-сервисов; численность населения; интенсивность интернет-дискуссии в зависимости от насыщенности политическими, экономическими или социальными событиями; цитирование страны глобальной интернет-аудиторией в зависимости от интереса к стране на глобальном рынке и политической арене; уровень обсуждения тех или иных технологий в связи с наличием этих технологий в стране или интересом к этим технологиям [Колесниченко и др., 2015; Колесниченко и др., 2016; Kolesnichenko et al., 2016].

Социологические опросы могут исказить истинную картину, например, когда при очном общении люди скрывают свое истинное мнение, а также когда на их мнение оказывает влияние совокупность факторов. Люди могут отвечать на вопросы, исходя из той социальной роли или статуса, в которых их застал опрос (на-

пример, на работе может быть высказано одно мнение, а дома — другое). Любой социологический опрос уникален и не глобален, он имеет свои границы. Интернет в этом смысле невозможно загнать в четкие рамки, в виртуальной среде путаются социальные статусы; невозможно точно для всех установить гендерные и возрастные различия, если сведения указаны ложно; один человек может продуцировать несколько «цифровых личностей» (ботов) и при этом «вживаться в роль», что с точки зрения классической социологии неприемлемо.

В исследовании использован метод подсчета ключевых слов через API-доступ. Такое исследование можно назвать API-социологическим. В данном аспекте используются подходы Больших данных (big data), допускающие, что массив собираемых данных не идеален, может содержать неточности и неопределенности. Количество подсчитываемых слов может меняться во времени из-за динамичности самого интернета — постоянного изменения числа сайтов и характеристик индексации сайтов поисковой системой. Неточность, но максимально возможный охват явления — это одно из качеств Больших данных, которое принимается за правило для аналитики. Такой «общий замер» по максимально возможному объему интернета позволяет «втянуть» в общий контур все варианты неопределенностей, включая ангажированность при составлении новостей и иных текстов или генерацию комментариев искусственно созданными «цифровыми личностями», при этом главный тренд все равно будет отражать истинную реакцию на события. Это как реакция толпы, даже если локально где-то влияют какие-то искажающие факторы, общая реакция будет иметь определенную направленность.

Методами морфологической текстовой аналитики Больших данных можно оценивать лишь отражение интереса к политическим, социальным и экономическим процессам в информационной системе «People-to-IT», а не сами эти процессы напрямую. Необходимо отметить, что число ключевых слов, отражающих интерес к технологиям, можно адресовать к феномену связанности диффузии технологий с наступлением новой волны технологического уклада, описанному японским глобалистом Масаки Хироока (Masaki Hirooka). Ученый показал нарастание диффузии в экономику новейших технологий в преддверии наступления новой Шестой Кондратьевской волны (2020—2070-е гг.) [Hirooka, 2006; Korotayev et al., 2011]. Количество подсчитанных ключевых слов по технологиям косвенно отражает как диффузию этих технологий в экономики стран, так и повышенный к ним интерес без реальной диффузии.

Стоит подчеркнуть, что ведущую роль в аналитике Больших данных играют разные методы статистического анализа и визуализации. В данном исследовании поэтапно расширяется перечень применяемых методов количественного анализа и вариантов визуализации данных при поиске скрытых закономерностей.

Методика

В анализ вошли 100 стран из разных регионов мира. Регистрируемые годы — 2015 и 2016. Изъятие данных (data mining) проводилось в течение 2017 г. при помощи специально созданного программного обеспечения. Через API-доступ к поисковой системе Google осуществлялся запрос для определения встречаемости заданных (ключевых) слов в открытых для общего доступа текстовых данных

интернета. Методически такой распределенный по разным базам данных сбор ключевых слов можно рассматривать как использование Google в качестве неклассического суперкомпьютера в режиме «as-a-Service». Языки набора слов — английский и русский (доля русских слов существенно меньше, чем английских).

Для анализа были выбраны 52 ключевых слова, разбитые на четыре блока (см. табл. 1). Найдены значения встречаемости каждого слова для всех стран и соответствующего года. Полученные данные для каждого года сформированы в матрицы размером 100×52. Визуализация табличных метаданных выполнена в Microsoft Excel. Осуществлен кластерный анализ стран методом k-means с использованием евклидовой метрики в интерактивной среде iPython (библиотеки NumPy, Pandas и Sklearn). Этот метод позволяет разбить выборку на группы с минимальным разбросом значений. С использованием методов полной связи и одиночной связи построены дендрограммы для переменных «страны» и «слова». Дендрографический анализ отображает взаимные связи между объектами из заданного множества.

Таблица 1. *Морфологическая матрица из 52 ключевых слов по data mining*

Тематические блоки ключевых слов

Unit «Information»

computer, cloud computing, mobile phone, dollar exchange rate, euro exchange rate, mobile app, software, artificial intelligence, smartphone

Unit «Energy & Material»

electric cars, solar panel, drip irrigation, gas supplies, oil, oil price, nuclear power plant, 3D-printing, robotics, GMO, GMO harmful

Unit «Social inequality»

crisis, inflation, price increase, unemployment, dismissal, poverty

Unit «Stress load»

migrants, refugees, Muslims, terrorism, terrorist, occupation, narcotic, alcoholism, morbidity, mortality, violation, crime, war, casualties

Unit «Social profile»

revolution, demonstration, protest, strike, corruption, stability, prosperity, democracy, development, freedom, human rights, justice

Для каждой страны построены лепестковые диаграммы, показывающие группы самых больших паттернов из списка ключевых слов (см. Приложение 1). На их основе найдено распределение стран по самому часто встречающемуся (ведущему) ключевому слову (см. рис. 1), что ранее в исследовании обозначалось как «имидж» государства — некая обобщенная характеристика по ведущему слову. Позитивный имидж — ведущий паттерн с конструктивным смыслом, негативный имидж — ведущий паттерн с деструктивным смыслом.

На основании лепестковых диаграмм, показывающих группы паттернов наиболее часто встречающихся ключевых слов для каждой из стран, была построена общая диаграмма, отражающая глобально самые актуальные темы, связанные со странами.

Сделан тематический анализ слов с применением трех потенциалов развития стран, предложенных научным руководителем Института экономики РАН Р. С. Гринбергом на X Международной Кондратьевской конференции в рамках Международного конгресса по глобалистике (2017 г.).

Основные результаты и обсуждение

Определение ведущих паттернов

Спектр ведущих паттернов (самое часто встречающееся ключевое слово в привязке к названию страны из списка 52 слов) составил 15 позиций (рис. 1). К позитивному имиджу можно отнести такие паттерны, как «development», «mobile app», «mobile phone». По полученным данным, только 32 страны (то есть 32 % стран) относятся к группе позитивного имиджа. Подгруппа «development» содержит 20 стран, это ведущие страны мира (Россия входит в эту подгруппу). Остальные страны попали в группу негативного имиджа с ведущими паттернами: «revolution», «protest», «strike», «demonstration», «poverty», «morbidity», «narcotic», «occupation», «Muslims», «terrorism», «war», «oil».

Ключевое слово «oil» (нефть) отнесено к негативному имиджу, так как оно показывает отставание в принятии экономикой страны технологического уклада информационных технологий. Ключевое слово «revolution», исходя из списка 20 стран, попавших в подгруппу с таким ведущим паттерном, отражает не технологическую революцию, а социальные процессы и настроения. Слово «occupation» изначально задумывалось как отражение темы оккупации территорий, но по анализу полученных результатов стало понятно, что оно также указывает и на проблемы трудовой занятости.

Обобщая данные, представленные на рис. 1, можно отметить, что в подавляющем большинстве случаев для стран самым актуальным трендом является социальная сфера, а не терроризм, война или технологии. Это дает понимание того, что в грядущие 2020-е годы, когда ожидается начало шестой Кондратьевской волны с новым технологическим укладом и с глобальной цифровой экономикой, мир будет подвергаться серьезным социальным волнениям, связанным с дезадаптацией к новым экономическим условиям.

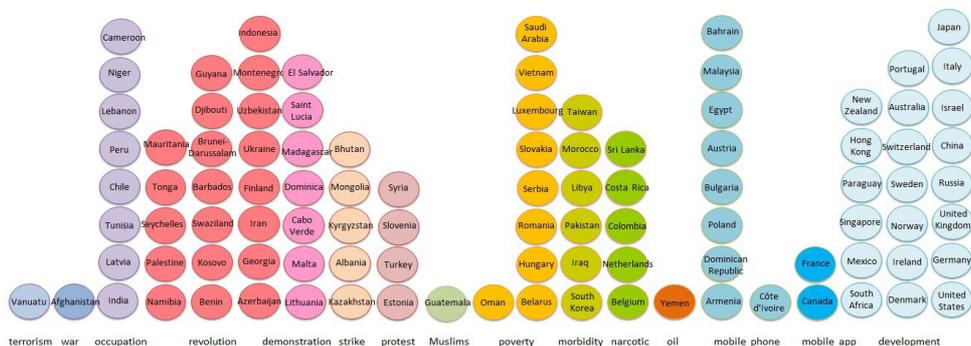
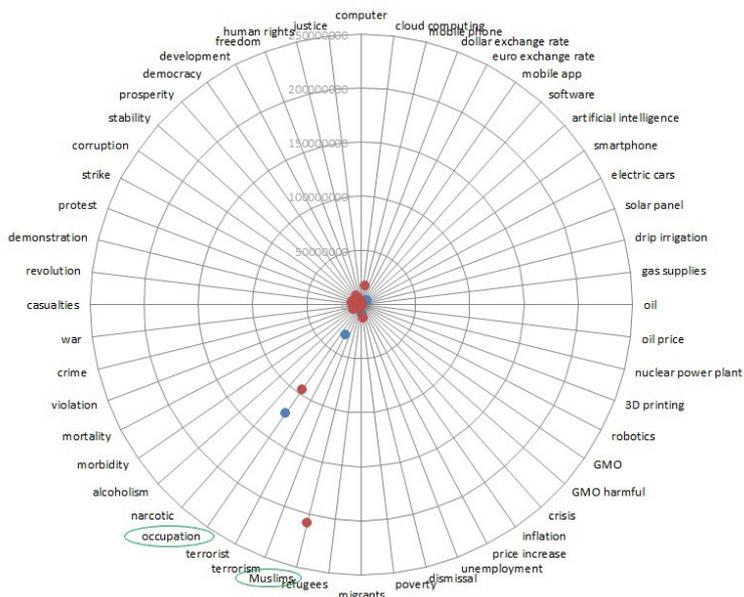
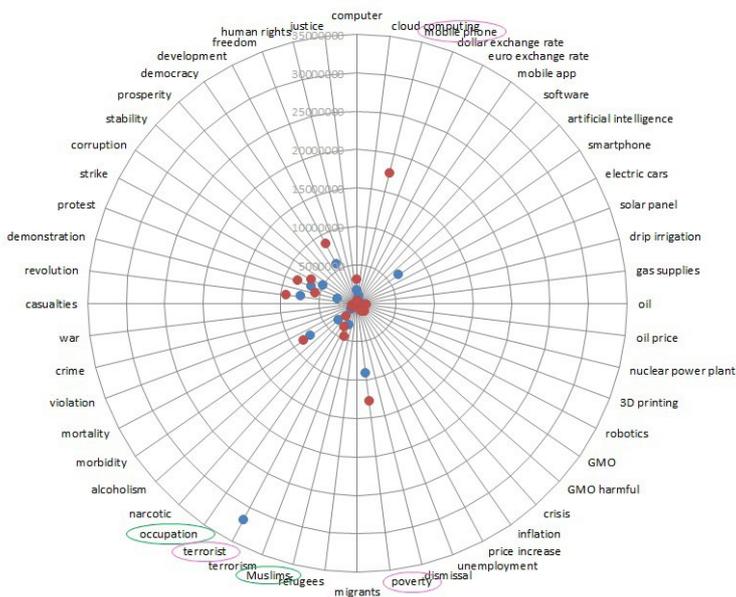


Рисунок 1. Распределение 100 стран по ведущему паттерну из 52 ключевых слов



А



Б

Рисунок 2. Диаграммы распределения паттернов ключевых слов, привязанных к названию Гватемалы за 2015—2016 гг. А — первый ряд больших паттернов; Б — второй ряд больших паттернов

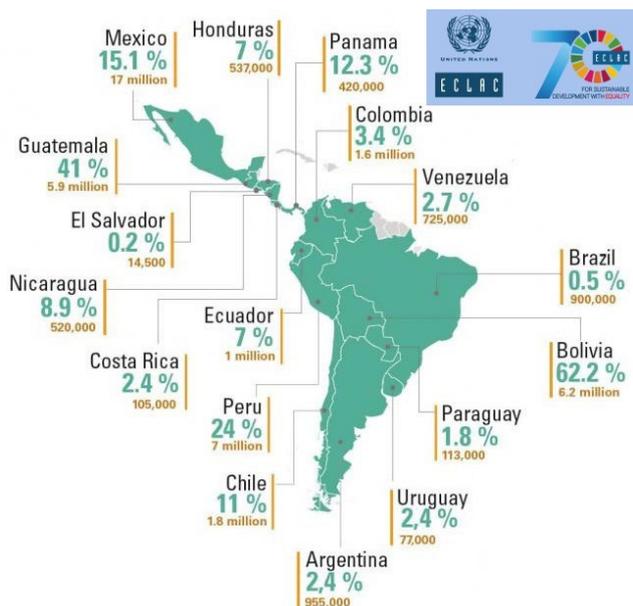


Рисунок 3. Процентное распределение бедных группы коренных индейцев в странах Латинской Америки 1



Рисунок 4. Фотоматериалы из блога «[isisandislaminlatinamericablog](https://isisandislaminlatinamericablog.wordpress.com/)»², надпись на фото слева сверху — «The Mayan Islamic Future 2020»

¹ Источник инфографики: Economic Commission for Latin America. URL: <https://www.cepal.org/en/infografias/los-pueblos-indigenas-en-america-latina> (accessed 07.03.2019).

² URL: <https://isisandislaminlatinamericablog.wordpress.com/> (на момент выхода статьи сайт заблокирован)

Внимание исследователей привлекли данные по Гватемале, оказавшейся единственной в подгруппе ведущего ключевого слова «Muslims» (мусульмане) на рис. 1. Первичные лепестковые диаграммы для Гватемалы представлены на рис. 2. По данным Экономической комиссии ООН, в странах Латинской Америки и Карибского бассейна насчитывается около 45 млн представителей коренных индейских народов, что составляет 8,3% от популяции региона³. Наибольший процент индейского населения, бедного и плохо интегрированного в некоренную испанскую преобладающую популяцию, встречается в Гватемале и Боливии (см. рис. 3). Мусульманское население в Гватемале крайне малочисленно, официально оно не превышает 1,5 тыс. человек. В основном это выходцы из Палестины⁴. Проблема социального беспокойства, отраженная в ведущем паттерне «Muslims», не может быть связана с потоком мигрантов-мусульман, так как его нет.

По данным некоммерческой организации USA for UNHCR (United Nations High Commissioner for Refugees, Агентство ООН по делам беженцев)⁵, наиболее актуальной современной проблемой Гватемалы стала организованная преступность, которую называют «maras». Это новое поколение бандформирований, пришедшее на смену «pandillas», бандам времен гватемальской гражданской войны 1960—1996 гг. «Maras» отличаются транснациональными корнями и тесно связаны с мигрантами [Does, 2013], число членов бандформирований превышает 22 тыс. человек⁶. В интернете можно найти информацию, что в преступном мире Гватемалы усиленно работают мусульмане-рекрутеры за-прещенной в России группировки ИГИЛ⁷, вовлекающие индейцев в свои ряды. Например, на сайте ISISandIslamInLatinAmericaBlog⁸ указано, что в Гватемале живут 23 разделенные этнические общины индейцев майя, сегрегированных от испанского большинства и прошедших массовый геноцид во времена гражданской войны. Данный сайт полон информации и фотографий (см. рис. 4), показывающих вовлеченность террористических исламских группировок в индейское население Гватемалы и граничащего с ней мексиканского штата Чьяпас (Chiapas), который населен преимущественно сегрегированными индейцами. Идеология мусульманского рекрутирования бедного индейского контингента связана с призывами отомстить за угнетение и оккупацию родных земель, при этом исламские международные террористы выступают в роли «справедливой силы мстителей», поднимающей народ на «праведный» бой (тренд обозначен как «The Mayan Islamic Future»).

³ Latin America Has Achieved Progress in Health, Education and Political Participation of Indigenous Peoples in the Last Decade. 2014 URL: <https://www.cepal.org/en/pressreleases/latin-america-has-achieved-progress-health-education-and-political-participation> (дата обращения: 19.04.2019)

⁴ Если полагаться на данные Википедии: «En. Wikipedia.org/wiki/Islam_in_Guatemala». Эти данные также подтверждаются отчетом Pew Research Center, выдержки из которого со статистическими данными распределения мусульман по странам мира, опубликован в газете The Guardian URL: <https://www.theguardian.com/news/datablog/2009/oct/08/muslim-population-islam-religion> (дата обращения: 19.04.2019).

⁵ URL: <https://www.unrefugees.org/refugee-facts/statistics/> (дата обращения: 19.04.2019).

⁶ Issue L. Faith-Based Organizations and Migrant Reintegration in Central America's Northern Triangle. U.S. Department of State, Office of Religion and Global Affairs. 2017.

⁷ ISISandIslamInLatinAmericaBlog.wordpress.com (на момент выхода статьи сайт заблокирован).

⁸ URL: <https://isisandislaminalatinamericablog.wordpress.com/> (на момент выхода статьи сайт заблокирован).

На рис. 2 видно, что в целом ведущими паттернами ключевых слов для Гватемалы являются «Muslims», «occupation», «terrorist», «poverty», что и отражает описанные выше процессы внутри страны. Пик паттерна ключевого слова «mobile phone» в совокупности с другими выявленными ведущими паттернами можно рассматривать как указание на активность террористических группировок, что ранее было подробно рассмотрено в нашей предыдущей работе, посвященной провокации экстремизма и терроризма через мобильную связь [Колесниченко и др., 2015].

На основании лепестковых диаграмм для каждой из стран, показывающих не одно ведущее слово, а группы часто встречающихся ключевых слов, в нашем исследовании была построена общая диаграмма для пиков паттернов первого уровня (см. рис. 5), которую мы назвали *картой глобальных трендов*. Также создана диаграмма для пиков паттернов второго уровня, без влияния более часто встречающихся слов первого уровня (рис. 6). В результате стало возможным развернуто в виде карты представить отражение глобальной дискуссии в интернете в стиле ежегодных докладов экспертов Всемирного экономического форума «The Global Risks Report» (GRR).

GRR создается на основе соцопросов экспертов разных стран. В докладе 2018 г. глубокая социальная нестабильность (profound social instability) занимает центральное место среди глобальных проблем и рисков (по ссылке можно ознакомиться с диаграммой The Risks-Trends Interconnections Map⁹). В докладах GRR от 2015—2017 гг. социальная нестабильность соотносилась по силе влияния с другими трендами — безработицей, миграцией, недостатками в госуправлении. А уже к 2018 г. социальная нестабильность сформировалась как лидирующий по влиянию глобальный тренд, тесно связанный с социальным неравенством, поляризацией общества, негативными последствиями внедрения новых технологий и увеличением доли пожилого населения. По динамике GRR с 2015 г. по 2019 г. на фоне роста влияния социальной нестабильности снизилась даже актуальность темы развития межгосударственного конфликта в мире.

На рис. 5 представлена полученная в данном исследовании карта глобальных трендов по ключевым словам (пики паттернов первого уровня). Чем ближе к центру по шкале, тем больше стран имеют конкретное ключевое слово как часто встречающееся в привязке к своему названию. Размер маркера условно отражает эту же характеристику. На карте глобальных трендов находятся четыре лидирующих по распространенности позиции — «mobile phone», «development», «revolution», «poverty». На втором месте по актуальности — «morbidity», «narcotic», «occupation» (как трудовая занятость), а также протестная тема: «protest», «strike», «demonstration». Данное графическое представление частоты встречаемости ключевых слов позволяет увидеть, какие темы (из обозначенного посредством 52 ключевых слов контура потенциальных тем) наиболее часто упоминаются в текстах в интернете. Стоит отметить, что это понимание основано на анализе диаграмм паттернов ключевых слов для каждой из ста проанализированных стран в отдельности, по результатам анализа данные были обобщены в целостную картину. Такой

⁹ The Global Risks Report 2018. 13th Edition. World Economic Forum, Geneva. URL: <http://wef.ch/risks2018> (дата обращения: 25.02.2018).

подход позволяет выявить актуальные для каждой из стран тенденции и избежать описания случайных распределений частоты слов, если бы они подсчитывались без привязки к названию страны.

Получившаяся карта отражает интерес глобального социума к технологиям нового информационного уклада (наиболее распространенная технология на данном этапе — мобильная связь); указание на развитие говорит о том, что социум не утратил конструктивного пути, но есть обусловленные сменой уклада проблемы — бедность, занятость. Революционный тренд может иметь двойное толкование, это и технологическая революция (точнее, широко обсуждаемая *The Fourth Industrial Revolution*), и революционные настроения в странах. Последнему аккомпанируют часто встречающиеся слова, отражающие готовность к протестам, забастовкам и демонстрациям, а также периферийный ряд трендов со словами, отражающими проблемы кризиса, повышения цен и безработицы. Проблемы заболеваемости и наркомании, входя в центральное звено главных трендов, беспокоят глобальный социум намного больше, чем терроризм, война и преступность. Тема алкоголизма находится в периферийном ряду трендов.

Две группы слов, отражающие новые технологии и уходящие (вернее, меняющие свою роль) технологии, равны по весу. Это компьютерные мобильные технологии, программное обеспечение — новый уклад. И углеводороды — лидеры в период старого уклада. Углеводороды не теряют актуальности в новом укладе, так как они меняют роль и занимают нишу нефтехимии (материалы для новой промышленности, основанной на 3D-печати). Интерес к ГМО (генетически модифицированным организмам) и оценка вреда генетических технологий также являются объективными трендами времени.

На рис. 6. представлена карта глобальных трендов по ключевым словам второго уровня, без влияния пиков паттернов первого уровня (то есть данные, отраженные на рис. 5, не входят в диаграмму на рис. 6). Принцип построения графика тот же: чем ближе к центру по шкале, тем больше стран имеют конкретное ключевое слово как часто (не редко) встречающееся в привязке к своему названию. Размер маркера условно отражает эту же характеристику. Общая картина более насыщена ключевыми словами в сравнении с рис. 5, так как это не самые пиковые группы трендов, а широкий пласт постоянно обсуждаемых проблем. Если для части стран не выпали обозначенные на рис. 5 ключевые слова как главные тренды первого уровня, то они все равно заполняют второй уровень глобальной дискуссии. То есть построенная *модель глобальной дискуссии* в принципе одинакова для обоих уровней пиков паттернов слов — и для самых часто обсуждаемых тем, и для менее топовых, но все же частых, базовых тем обсуждения в текстах. Можно лишь отметить, что на этом базовом уровне в сравнении с рис. 5 более весомое место занимает тема терроризма. Построение карты с пиками паттернов второго уровня позволяет не пропустить важные тенденции, которые могут быть скрыты за численно выраженными пиками первого уровня.

Сравнивая с картами GRR последних лет оба рис. 5 и 6, необходимо подчеркнуть принципиальную схожесть главных трендов, а именно: лидируют социальные проблемы, указывающие на социальную нестабильность, обусловленную текущей сменой технологического уклада.

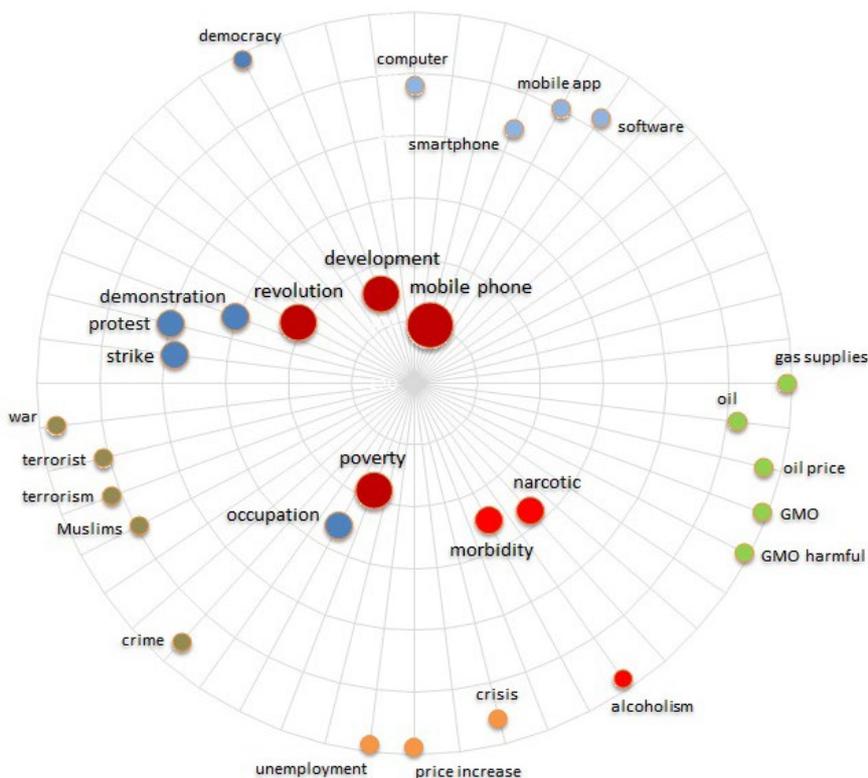


Рисунок 5. Карта глобальных трендов, построенная по группе самых часто встречающихся паттернов ключевых слов для каждой из 100 стран; чем ближе к центру по шкале, тем больше стран имеют конкретное ключевое слово как часто встречающееся в привязке к своему названию (размер маркера условно отражает эту же характеристику)

Если принять за условие, что построенная модель глобальной дискуссии отражает тот образ мира, в котором социум сегодня существует, то можно охарактеризовать этот глобальный образ. Понимание этого образа дает возможность более точно прогнозировать процессы глобального развития и оценить риски будущего. Например, без такого понимания общей картины вовлеченным в принятие решений экспертам может субъективно казаться, что мировая дискуссия переполняется призывами к войне, или же что все обсуждают только динамику цен на нефть. А на самом деле картина совершенно иная. Ошибочное субъективное восприятие ситуации обуславливает неправильное направление прогноза развития событий, а неправильный прогноз, в свою очередь, приводит к ошибочным действиям. Наша методика не претендует на абсолютную точность оценки глобальной ситуации, однако она дает более конкретную картину в сравнении с глобальными оценками и прогнозами, построенными на мнении фокус-групп и опросов экспертов.

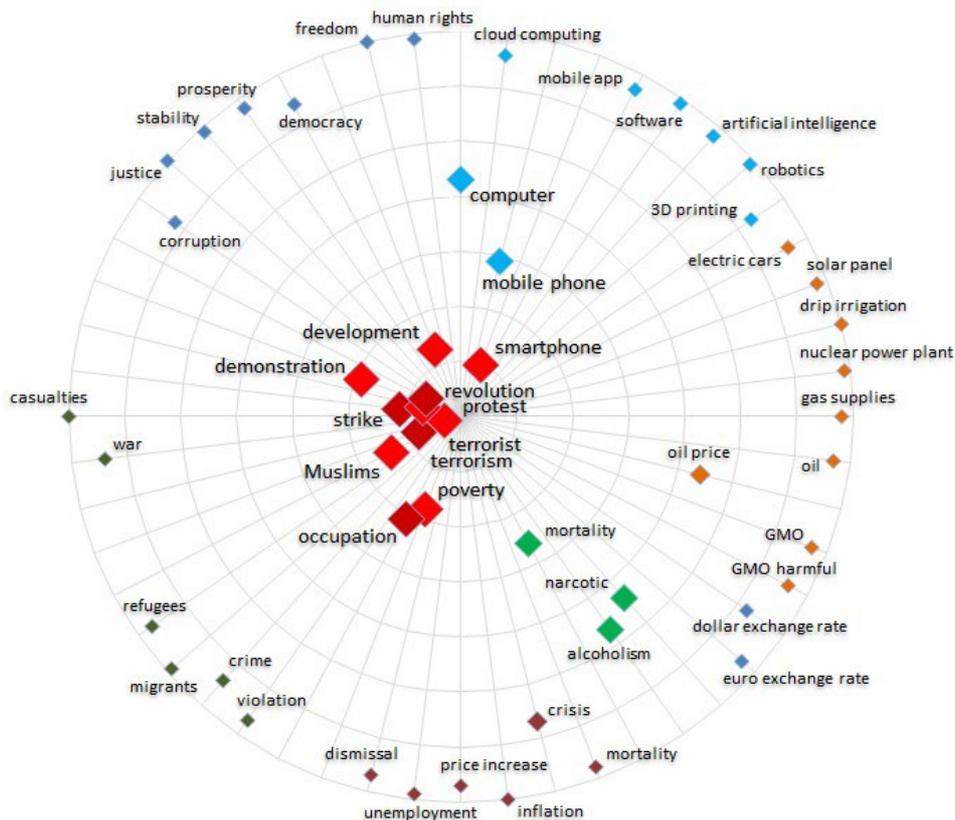


Рисунок 6. Карта глобальных трендов второго уровня, построенная без влияния пиков паттернов ключевых слов первого уровня для каждой из 100 стран; чем ближе к центру по шкале, тем больше стран имеют конкретное ключевое слово как часто встречающееся в привязке к своему названию (размер маркера условно отражает эту же характеристику)

В аспекте сравнения интересен международный доклад MFP-2030 (Multiple Futures Project — Navigating Towards 2030), сделанный в 2009 г.¹⁰ Экономисты и военные вместе попытались предвидеть будущие риски для развития глобального социума, описав не единую картину, а четыре возможных сценария развития глобальных событий. MFP-2030 ранее был подробно рассмотрен в серии статей в газете «Военно-промышленный курьер» в 2009 г.¹¹

Первый сценарий по MFP-2030 — темная сторона эксклюзивности (dark side of exclusivity): слабые, социально нестабильные государства, рост радикальных идей и национализма, усиливающееся демографическое давление с Востока, расширение и углубление зоны бедности, неконтролируемая миграция. Второй сце-

¹⁰ Multiple Futures Project — Navigating towards 2030. Report. ACT NATO. 2009. 69 p.

¹¹ См., например: Юрєва О. (Колесниченко О. Ю.). Один мир, четыре сценария будущего // Газета «Военно-промышленный курьер». 2009. 33 (299). С. 2.

нарий — обманчивая стабильность (*deceptive stability*): акценты делаются на трансграничной преступности, терроризме и экстремизме, угрозах из регионов бедных стран. Третий сценарий — столкновение современных векторов развития (*clash of modernities*): «гремучий коктейль» из технологических инноваций, интеллектуального пиратства, кибертерроризма, попадания высоких технологий в руки террористов и криминальных элементов, диссонанса между технологическими достижениями и гражданскими свободами; усиление конкуренции за энергию, нефть, газ, металлы и минералы, воду, пищу. Четвертый сценарий — новые политические силы (*new power politics*); усиление региональных противостояний в мультиполярном мире, военные действия, межгосударственные конфликты.

Очевидно, что прогноз MFP-2030 нельзя отнести к точному, это больше похоже на осознание экспертами, принявшими участие в опросах, несостоятельности субъективного подхода применительно к задачам глобальных оценок и прогнозирования. Если попытаться найти общие черты между построенной моделью глобальной дискуссии в нашем исследовании и четырьмя сценариями будущего по MFP-2030, то можно сказать, что на данном этапе нет указаний на развитие четвертого сценария, также не просматривается четко и третий сценарий. Два этих сценария наиболее агрессивны, с усилением военных действий и терроризма. Скорее всего, более вероятен первый сценарий или среднее между первым и вторым. Ведь именно первому сценарию свойственно ослабление государств из-за внутренних проблем (бедность, безработица, болезни и наркотики), на что и указывает построенная модель глобальной дискуссии интернет-аудитории.

Если упрощенно обозначить перечисленные сценарии будущего, то мир может трансформироваться в четыре модели развития: социальную, криминальную, технологическую и военную. Пока что по данным нашего исследования глобальная дискуссия сохраняется в рамках социальной модели. Само название первого сценария подводит к мысли, что проблемы нового технологического уклада будет невозможно решать государствам в одиночку, поддерживая свою эксклюзивность. Необходимы международные согласованные меры, релевантные социальной и экономической глобализации и происходящей трансформации. Однако до 2030 г. еще есть время, и, возможно, при условии усугубления дезадаптации к новому укладу мировой ландшафт поменяется в направлении третьего или четвертого сценария.

Сравнение потенциалов стран

На V Международном конгрессе по глобалистике (Москва, МГУ им. М. В. Ломоносова, 25—30 сентября 2017 г.), на прошедшей в рамках конгресса X Международной Кондратьевской Конференции президент Международного Фонда Н. Д. Кондратьева Руслан Семенович Гринберг очертил три сферы конкурентоспособного преимущества России на мировом рынке — три основных потенциала страны¹².

Первый потенциал — природный, природные богатства земли и морей России, делающие страну самой богатой территорией мира.

¹² Гринберг Р. С. Стенограмма выступлений президиума конференции. X Международная Кондратьевская конференция «Научное наследие Н. Д. Кондратьева и современность», V Международный Научный Конгресс «Глобалистика». Москва: МГУ им. М. В. Ломоносова. 2017.

Второй потенциал — интеллектуальный, обусловленный высоким уровнем образовательных традиций и талантом населения, включая молодое поколение.

Третий потенциал — пространственный, территориальный, позволяющий развивать масштабные инфраструктуры и транспорт.

Для каждого из этих потенциалов Р.С. Гринберг обозначил тормозящие факторы, которые могут полностью нивелировать все потенциальные возможности и остановить рост экономики. Для природного потенциала это увеличивающееся неравенство среди населения, бедность. Для интеллектуального потенциала — снижение уровня образования молодежи. Для пространственного потенциала — плохая транспортная инфраструктура и неконтролируемые потоки мигрантов.

В нашем исследовании проведен анализ этих трех потенциалов в сравнении с несколькими странами. В основу графиков на рис. 7, 8, 9 положено количество встречаемости ключевых слов, оси тематически отражают потенциалы стран, а размер маркеров — тормозящие факторы. Данный тематический вариант текстовой морфологической аналитики оказался очень информативным.



Рисунок 7. Отражение природного потенциала стран и тормозящего фактора



Рисунок 8. Отражение интеллектуального потенциала стран и тормозящего фактора

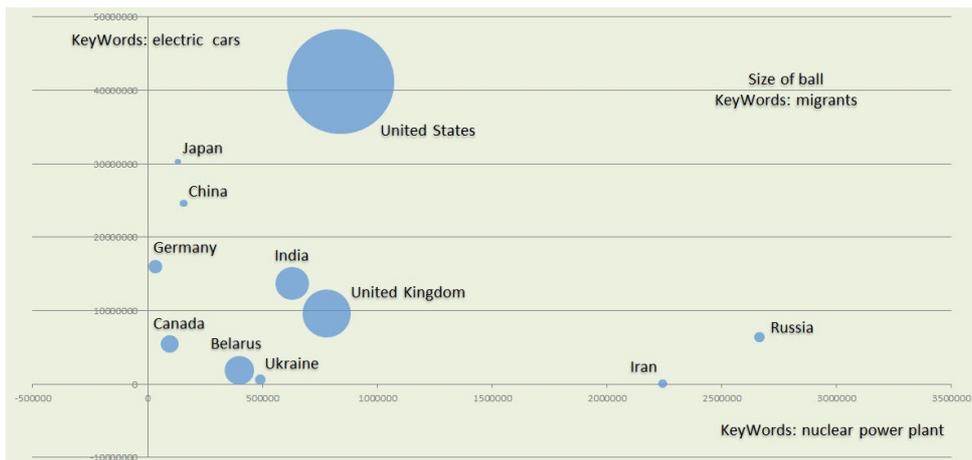


Рисунок 9. Отражение пространственного потенциала стран и тормозящего фактора

На рис. 7 можно заметить, что обсуждение проблем бедности в привязке к России сопоставимо с другими представленными странами, за исключением Украины, Беларуси и Ирана, для которых ключевое слово «poverty» встречается чаще в два и более раз. Данный график также отразил снижение интереса к нефти со стороны Великобритании, России, Японии, США, при сохранении более высокого интереса у Германии, Канады, Индии, Китая. Низкий интерес к теме нефти характерен для стран, успешно перестраивающихся к условиям нового технологического уклада.

На рис. 8 можно увидеть, что в привязке к России (как, впрочем, и к Великобритании) реже встречается ключевое слово «software», отражающее важное направление экономической независимости в новых условиях цифровой экономики. В то время как США — абсолютный лидер по интересу к технологиям нового технологического уклада. В качестве указания на тормозящий фактор выбрано слово «narcotic». Из представленных на графике стран худший показатель у Германии, для России тема наркомании тоже актуальна.

На рис. 9 заметно, насколько на США, лидирующие в развитии нового вида транспорта, оказывает негативное влияние проблема миграции. Для России отмечен высокий интерес к ядерной энергетике, что с точки зрения мирового перехода на электротранспорт весьма выгодно, так как создает устойчивую инфраструктуру для генерации и поставок электроэнергии для электротранспорта.

Таким образом, в рамках нашего исследования по API-социологии в сравнении с рядом стран Россия по трем потенциалам занимает скорее устойчивые и перспективные позиции, чем проигрышные, с поправкой на необходимость корректировать такие явления, как наркомания, слабый отечественный рынок программного обеспечения и пока еще недостаточное внимание к отечественному электроавтомобилестроению.

Дендрографический анализ

Впервые в данном исследовании по API-социологии, проводимом с 2015 г., представлены результаты дендрографического анализа (разновидность кластер-

ного анализа). Для матрицы слов по 2015 г. и матрицы по 2016 г. получен одинаковый результат в отношении США — эта страна стоит отдельно от всех остальных выбранных стран. Учитывая набор ключевых слов, а также самую интенсивную активность США в интернете, данный факт свидетельствует о технологическом лидерстве США. Указания на это были получены и в ходе предыдущего этапа исследования.

Дендрографический анализ (иерархическая кластеризация) выполнен двумя методами — одиночной связи и полной связи. Для определения расстояния между объектами (или кластерами) используется разная мера. В случае метода одиночной связи (Single Linkage) идет поиск двух наиболее близких объектов — это правило «ближайшего соседа» (Nearest Neighbor), далее к паре присоединяются расположенные ближе следующие объекты. Этот метод лучше подходит для выявления ограниченного числа крупных кластеров. Если за меру принимается как можно более удаленное положение объектов, то это метод полной связи (Complete Linkage) или «дальнего соседа» (Furthest Neighbor). Объект присоединяется к кластеру, если самый далекий элемент этого кластера находится ближе всего к новому объекту, чем самые далекие элементы других кластеров. Метод полной связи позволяет обнаружить большее число мелких кластеров.

На рис. 10 представлен фрагмент дендрограммы (матрица за 2016 г., метод полной связи), на котором видно, что США находятся отдельно от всех стран, а Россия имеет наиболее близкое родство с Великобританией (полностью дендрограмма представлена в *Приложении 2*). На дендрограмме по матрице за 2015 г. Россия имеет близкое родство с Австралией. Результат говорит о том, что в целом особенности распределения частоты ключевых слов в заданном множестве данных по России схожи со странами англосаксонской культуры.

На рис. 11 представлен фрагмент дендрограммы (матрица за 2016 г., метод одиночной связи), где анализируемыми параметрами стали ключевые слова (полностью дендрограмма приведена в *Приложении 3*). Этот фрагмент интересен распределением слов на два тематических семейства. Слева группа слов отражает темы военных действий и преступности, и к этим словам попало слово «smartphone», что может свидетельствовать о роли мобильной интернет-связи в этих двух сферах. Группа слов справа объединила нефть и новые тренды технологического развития, включая 3D-печать. Такая связь характерна и показательна для формирования нового типа промышленного производства, где продукты нефтехимии используются в качестве «чернил» для 3D-печати. Один из будущих масштабных рынков промышленной 3D-печати — производство электромобилей (full-body 3D-printed car из армированного углеродным волокном термопластика / carbon-fiber reinforced plastic, CFRP). Электромобили для управления имеют «начинку» из алгоритмов искусственного интеллекта, связанных с миром интернета вещей через облачные технологии и мобильные приложения. Для зарядки электромобилей в США строят станции с солнечными батареями. Родство всех этих тем отражено в крупном кластере на полученной дендрограмме.

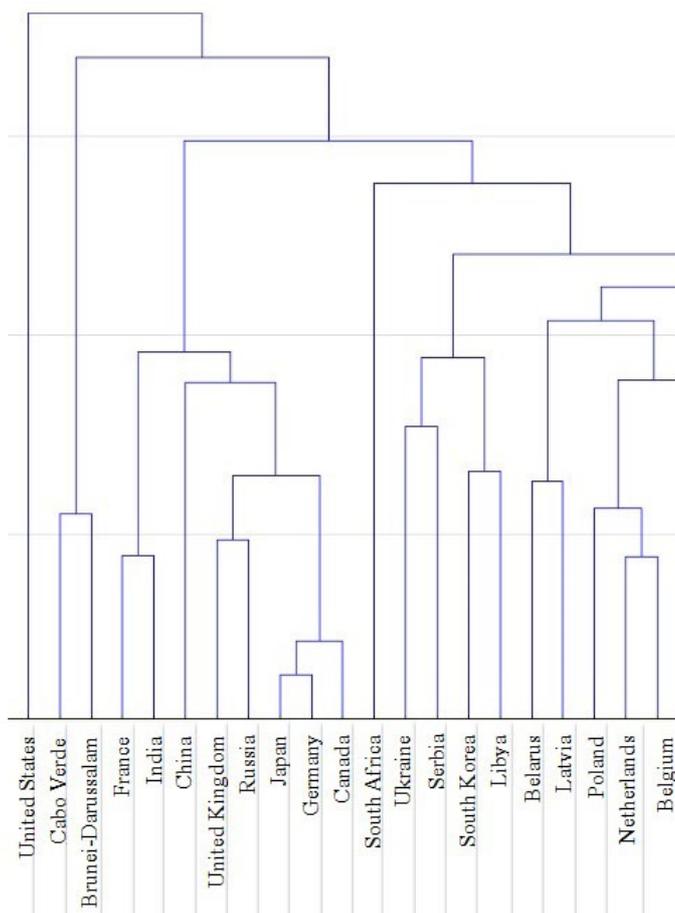


Рисунок 10. Фрагмент дендрогаммы по матрице из 52 ключевых слов и 100 стран, 2016 г.; анализируемые параметры — страны

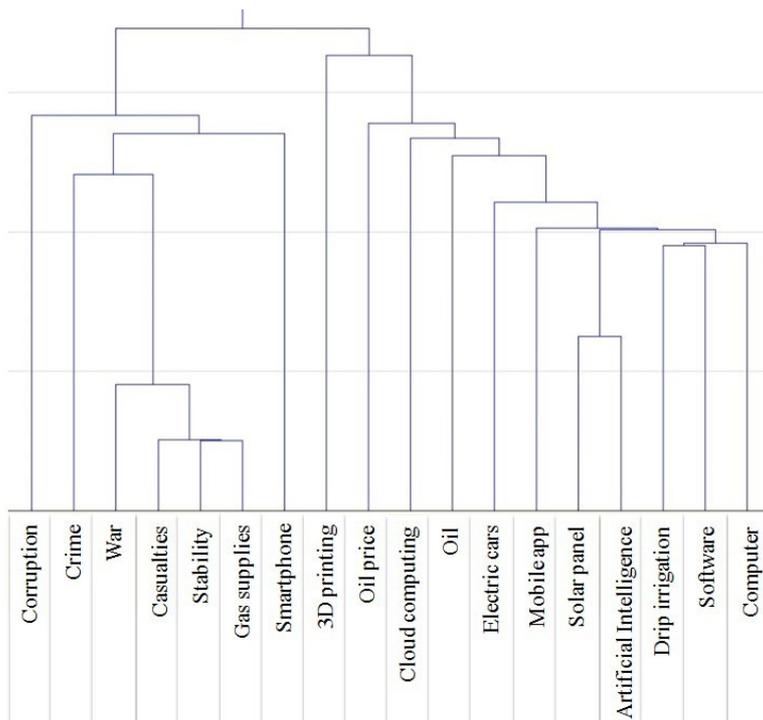


Рисунок 11. Фрагмент дендрограммы по матрице из 52 ключевых слов и 100 стран, 2016 г.; анализируемые параметры — ключевые слова

Кластерный анализ

В данной статье кластерный анализ не представлен подробно. Излагая результаты кратко, необходимо отметить, что по переменной «страна» наилучшей кластеризацией является разбиение на шесть групп. Распределение на кластеры в основном происходит по принадлежности стран к высокой, средней или низкой активности в интернете, однако на формирование кластера могут оказать влияние и особенности в распределении паттернов ключевых слов внутри матрицы. С этой точки зрения из полученных кластеров привлек внимание кластер за 2015 г., в который попало восемь стран: Йемен, Южная Корея, Бельгия, Нидерланды, Польша, Латвия, Беларусь и Казахстан. Интересно, что в матрице за следующий 2016 г. такого кластера не выявлено, но все эти восемь стран имеют в 2016 г. слово «narcotic» в качестве часто встречающегося слова первого и второго уровня.

Стоит отметить, что в целом в данном исследовании слово «narcotic» как ведущий паттерн отмечено у пяти стран (см. рис. 1). Среди часто встречающихся ключевых слов первого уровня слово «narcotic» имеют 25 стран, а среди частых слов второго уровня — 26. Таким образом, всего 51 страна из 100 (то есть 51 %) имеет слово «narcotic» в качестве часто встречающегося слова первого и второго уровня (см. Приложение 4).

Что объединило восемь стран в заинтересовавший исследователей кластер, можно понять, посмотрев табл. 2. Это совпадения ключевых слов в качестве часто встречающихся преимущественно негативного и дестабилизирующего характера. Ключевое слово «mobile phone» указывает на роль связанности общества через персональную мобильную связь, которая усиливает любые внутренние процессы, от протестных настроений до терроризма и преступности. Слова «development» и «democracy» хотя и относятся к позитивной группе слов, но в данном контексте лишь указывают на социальную значимость выявленных особенностей кластера.

На рис. 12 представлена карта мира, на которой маркерами-ромбами отмечены страны, характеризующиеся частой встречаемостью ключевого слова «narcotic» (первый и второй уровень паттернов частых слов). Маркеры совпадают с реальными маршрутами афганского героинового наркотрафика — балканским и северным, а также с маршрутом кокаина в США из Колумбии¹³. Маркер-звездочка показывает восемь стран из обсуждаемого кластера. Математическая аналитика Больших данных выходит за грань человеческих возможностей увидеть закономерности, за грань интуиции исследователя, о чем часто говорят ученые, представляющие те или иные современные методы математической обработки параметров. В данном случае в кластер страны объединились за год до того, как в их информационном поле стало доминировать слово «narcotic». Можно предположить, что эти восемь стран подверглись каким-то похожим глубинным негативным социальным процессам (что нашло отражение в соотношении паттернов ключевых слов в рамках единой матрицы), на фоне которых наркомания приобретает особое разрушительное воздействие.



Рисунок 12. Карта (ресурс Google) с нанесенными маркерами-ромбами, указывающими на страны, для которых ключевое слово «narcotic» является часто встречающимся (первый и второй уровень паттернов частых слов). Маркеры-звездочки указывают на страны из обсуждаемого в тексте кластера. Маркеры совпадают с реальными маршрутами афганского героинового наркотрафика — балканским и северным, а также с маршрутом кокаина в США из Колумбии

¹³ World Drug Report 2017. UNODC, Vienna, 2017. URL: <https://www.unodc.org/wdr2017/index.html> (дата обращения: 25.02.2018).

Таблица 2. **Совпадение ключевых слов для восьми стран кластера* в качестве паттернов часто встречающихся слов за 2015—2016 гг.**

Совпадение 100 % (8 стран)	Совпадение 75 % (6 стран)	Совпадение 50 % (4 страны)
narcotic	protest	Muslims
poverty	demonstration	terrorism
strike	development	mobile phone
revolution	corruption	morbidity
	occupation	GMO harmful
	crime	democracy
	alcoholism	
	terrorist	

* Йемен, Южная Корея, Бельгия, Нидерланды, Польша, Латвия, Беларусь и Казахстан.

Заключение

Проблема анализа интернет-данных обсуждается социологами уже несколько лет. Активной профессиональной площадкой для таких дискуссий традиционно выступает ежегодная Грушинская конференция¹⁴. После первого этапа сомнений о пригодности Больших данных для генерации научно-практических выводов наступил следующий этап проб и ошибок, когда возникла масса исследований с разными целями и методами. По мере развития интернет-социологии определились два направления — сбор IP-данных (Internet Protocol) и сбор данных через API. Сейчас уже можно расставить некоторые акценты.

Во-первых, крупные агрегаторы данных, предоставляющие платформы для разного рода взаимодействия людей в виртуальной среде, такие как Google, Facebook, Twitter, «ВКонтакте», «Яндекс» и т. д., имеют свой уникальный дизайн. Именно эта виртуальная инфраструктура и представляет научный интерес, не нужно дополнительно придумывать что-то новое. Важно исследовать то, как люди осуществляют взаимодействие на популярных интернет-площадках, что происходит, как это влияет на людей. В том или ином виде все крупные агрегаторы данных предоставляют возможность для аналитики, либо можно применять к ним дополнительно созданные инструменты для майнинга данных.

Во-вторых, не стоит «притягивать» классические социологические исследования к интернет-среде, так как интернет-социум — это иное явление, имеющее свои еще не познанные законы и тенденции. Определенно можно сказать, что человеческое общество подвергается благодаря интернету трансформации по трем направлениям: виртуализация жизни, хаотизация взаимосвязей и статусов, ускорение течения всех социальных процессов внутри связанного информационными технологиями социума. Конечно, исследования, в которых проводят интернет-соцопросы, анализируют выборку конкретных людей (стараясь отсекаль дубли-

¹⁴ Федоров В. В. Стенограмма выступлений президиума на пленарном заседании «Большая социология: интеграция новых данных и экспансия в междисциплинарность». V социологическая Грушинская конференция «Большая социология: расширение пространства данных». М.: РАНХиГС, 2015.

рующие виртуальные аккаунты или боты), делят на кластеры сегменты соцсетей, анализируют тематические группы, строят графы, отслеживают маршруты посещения сайтов, реакцию на события и публикации СМИ, комментарии и хэштеги и т. д. — уникальны и важны, они имеют нескончаемое число вариантов и уже стали поводом для описания новых тенденций в жизни общества. Не менее интересен и анализ текстов, причем не только Sentiment Analysis (анализ тональности текста), но и морфологический анализ, при котором подсчитываются ключевые слова и сравниваются паттерны частоты встречаемости определенных слов в текстах или запросах [Андрианов и др., 2014; Беляков, 2016а; 2016b; Варламов, Турдаков, 2016; Джапек, 2015; Посевкин, Бессмертный, 2015; Dietrich et al., 2015]. Ярким примером морфологического анализа служит сервис Google Trends поисковой системы Google¹⁵. Google-Тренды сами по себе интересны для анализа, но можно делать и свое исследование Google, применяя специально разрабатываемое программное обеспечение (что и сделано в представленном в данной статье исследовании). Особо выдающиеся работы по изучению интернета и мобильных сетей принадлежат исследователям в области Network Theory Альберту-Ласло Барабашу [Barabási, 2002; Barabási, 2016] и Марку Ньюману [Newman, 2010].

Резюмируя описанные результаты исследования, необходимо отметить следующее.

- Морфологический анализ с представленной конкретной матрицей ключевых слов позволяет изучить глобальный образ мировой интернет-дискуссии, отражающий процессы трансформации, происходящие в мире. Для каждой из стран можно оценить степень адаптации к меняющимся условиям технологического уклада и уровень социального стресса.
- Паттерны ключевых слов могут рассматриваться как информационные следы диффузии технологий в экономиках стран, с поправками на интенсивность общего интернет-присутствия страны или темы страны и возможность дискуссии по вопросу технологии, которой еще нет, но ее уже активно обсуждают.
- Дендрографический анализ, применяемый в генетических исследованиях, показал результативность в морфологической текстовой аналитике.
- Паттерны слов «narcotic» в привязке к названиям стран при наложении их на карту мира совпали с основными международными общеизвестными путями наркотрафика. А кластерный анализ выявил страны, в которых наркомания скорее всего имеет схожее системное негативное воздействие на жизнь социума, отражение чего определил математический алгоритм, найдя одинаковые взаимосвязи в рамках анализируемой матрицы.

Выводы

Результаты проведенного исследования по API-социологии позволили сделать следующие выводы:

1. Модель глобальной дискуссии, относящаяся к текстовым Большим данным Google за 2015—2016 гг., имеет социальную направленность и соотносится с трендами социальной нестабильности по данным ежегодного доклада «The Global Risks

¹⁵ URL: <https://trends.google.com/trends/> (дата обращения: 19.03.2014)

Report» Всемирного экономического форума. Построенная модель глобальной дискуссии релевантна одному из сценариев будущего, описанных в докладе MFP-2030, который называется «темная сторона эксклюзивности» (dark side of exclusivity). Для этого сценария характерно ослабление государств из-за внутренних социальных проблем, обусловленных сменой технологического уклада в глобальной экономике. Эти проблемы государства не могут решать в одиночку, поддерживая свою эксклюзивность. Необходимы международные согласованные меры по адаптации к социальной и экономической глобализации и трансформации.

2. Тематический анализ по трем потенциалам развития — природному, интеллектуальному, пространственному — показал, что в сравнении с рядом стран Россия занимает скорее устойчивые и перспективные позиции, чем проигрышные, с поправкой на необходимость корректировать такие явления, как наркомания, слабый рынок программного обеспечения и пока еще недостаточное внимание к отечественному электроавтомобилестроению.

3. По результатам дендрографического анализа Россия и Великобритания, а также Австралия имеют близкое родство, схожие особенности распределения частоты ключевых слов в заданном множестве данных за 2015 и 2016 гг. То есть отражение в интернете социально-экономических тем для России и стран с англосаксонской культурой имеет схожие черты.

4. Йемен, Южная Корея, Бельгия, Нидерланды, Польша, Латвия, Беларусь и Казахстан по результатам текстовой аналитики могут быть отнесены к странам с глубокими изменениями социально-экономической сферы из-за распространенности употребления наркотиков.

Список литературы (References)

Андрианов И. А., Турдаков Д. Ю., Астраханцев Н. А., Недумов Я. Р., Сысоев А. А., Майоров В. Д., Федоренко Д. Г., Кузнецов С. Д., Коршунов А. В. Texterra: инфраструктура для анализа текстов // Труды Института системного программирования РАН (электронный журнал). 2014. 1 (26). С. 421—438.

Andrianov I. A., Turdakov D. Yu., Astrakhantsev N. A., Nedumov Ya.R., Sysoev A. A., Mayorov V. D., Fedorenko D. G., Kuznetsov S. D., Korshunov A. V. (2014) Texterra: infrastructure for text analysis. In: *Proceedings of the Institute for System Programming of the Russian Academy of Sciences* (electronic journal). Vol. 1. No. 26. P. 421—438. (In Russ.)

Беляков М. В. Анализ новостных сообщений сайта МИД РФ методом контент-анализа (статья 1) // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2016а. № 3. С. 58—67.

Belyakov M. V. (2016a) The Analysis of News Messages of the Site of the Russian Federation Ministry of Foreign Affairs Applying Content-Analysis (Article 1). *RUDN Journal of Language Studies, Semiotics and Semantics*. No. 3. P. 58—67. (In Russ.)

Беляков М. В. Анализ новостных сообщений сайта МИД РФ методом сентимент-анализа (статья 2) // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2016b. № 4. С. 115—124.

Belyakov M. V. (2016b) The Analysis of News Messages of the Site of the Russian Federation Ministry of Foreign Affairs Applying Content-Analysis (Article 2). *RUDN Journal of Language Studies, Semiotics and Semantics*. No. 4. P. 115—124. (In Russ.).

Варламов М. И., Турдаков Д. Ю. Обзор методов извлечения информации из веб-ресурсов // Программирование. 2016. Т. 42. № 5. С. 30—48.

Varlamov M. I., Turdakov D. Yu. (2016) A survey of methods for the extraction of information from Web resources. *Programming and Computer Software*. Vol. 42. No. 5. P. 30—48. (In Russ.).

Джапек Л., Крейтер Ф., Берг М. и др. Отчет AAPOR о Больших данных: 2015 / Американская ассоциация исследователей общественного мнения ; пер. с англ. Д. Рогозина, А. Ипатовой, Е. Вьюговской ; предисловие Д. Рогозина. М., 2015.

Japex L., Kreuter F., Berg M., et al. (2015) Big Data in Survey Research: Aapor Task Force Report. Translated by D. Rogozin, A. Ipatova, E. Vyugovskaya. Foreword by D. Rogozin. Moscow. (In Russ.).

Колесниченко О. Ю., Смородин Г. Н., Ильин И. В., Журенков О. В., Мазелис Л. С., Яковлева Д. А., Дашонок В. Л. «Третья волна»: многоцентровое исследование по аналитике Big Data Академического партнерства EMC в России и СНГ // Мониторинг общественного мнения : Экономические и социальные перемены. 2015. № 5. С. 21—41.

Kolesnichenko O. Yu., Smorodin G. N., Ilyin I. V., Zhurenkov O. V., Mazelis L. S., Yakovleva D. A., Dashonok V. L. (2015) The Third Wave": Big Data Analytics Multicenter Study Provided by EMC Academic Alliance in Russia & CIS. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 21—41. (In Russ.).

Колесниченко О. Ю., Смородин Г. Н., Яковлева Д. А., Мазелис Л. С., Колесниченко Ю. Ю. API-социология: оценка глобальных деструктивных влияний на глокализационную ситуацию // Мониторинг общественного мнения : Экономические и социальные перемены. 2016. № 6. С. 30—51. <https://doi.org/10.14515/monitoring.2016.6.03>.

Kolesnichenko O. Yu., Smorodin G. N., Yakovleva D. A., Mazelis L. S., Kolesnichenko Yu. Yu. (2016) API in Sociology: Assessing Global Destructive Effects on Glocalization. *Monitoring of Public Opinion: Economic and Social Changes*. No. 6. P. 30—51. <https://doi.org/10.14515/monitoring.2016.6.03> (In Russ.).

Кондратьевские волны: наследие и современность / отв. ред. Л. Е. Гринин, А. В. Коротаев, В. М. Бондаренко. Волгоград : Учитель, 2015. [

Kondratiev waves: heritage and contemporaneity: Yearbook. (2015) Ed. by Grinin L. E., Korotaev A. V., Bondarenko V. M. Volgograd: Uchitel'. (In Russ.).

Посевкин Р. В., Бессмертный И. А. Применение сентимент-анализа текстов для оценки общественного мнения // Научно-технический вестник информационных технологий, механики и оптики. 2015. Vol. 1. No. 15. С. 169—171.

Posevkin R. V., Bessmertny I. A. (2015). Texts Sentiment-Analysis Application for Public Opinion Assessment. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. Vol. 1. No. 15. P. 169—171. (In Russ.).

Barabási A-L. (2016) *Network Science*. UK: Cambridge University Press.

Barabási A-L. (2002) *Linked: The New Science of Networks*. USA: Perseus Books Group.

Dietrich D., Heller B., Yang B. (2015) *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*, USA: Wiley.

Does A. (2013) *The Construction of the Maras. Between Politicization and Securitization*. Monograph. Graduate Institute Publications. <https://doi.org/10.4000/books.iheid.716>.

Hirooka M. (2006) *Innovation Dynamism and Economic Growth. A Nonlinear Perspective*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.

Kolesnichenko O., Smorodin G., Yakovleva D., Mazelis L., Zhurenkov O., Kolesnichenko Yu. (2016) Text Big Data Analytics: exploring API opportunity. In: *10th IEEE International Conference on Application of Information and Communication Technologies — AICT-2016*. Qafqaz University, Azerbaijan, Baku. P. 37—41.

Korotayev A., Zinkina Ju., Bogevolnov Ju. (2011) Kondratieff Waves in Global Invention Activity (1900—2008). *Technological Forecasting & Social Change*. Vol. 7. No. 78. P. 1280—1284. <https://doi.org/10.1016/j.techfore.2011.02.011>.

Newman M. E. J. (2010) *Networks. An Introduction*. UK: Oxford University Press.