



СЛОВО МОЛОДЫМ ПОЛИТОЛОГАМ

Известия Саратовского университета. Новая серия. Серия: Социология. Политология. 2022. Т. 22, вып. 2. С. 232–236

Izvestiya of Saratov University. Sociology. Politology, 2022, vol. 22, iss. 2, pp. 232–236

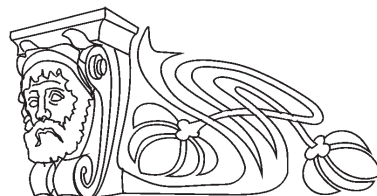
<https://soziopolit.sgu.ru>

<https://doi.org/10.18500/1818-9601-2022-22-2-232-236>

Научная статья

УДК 32.019.5+303.717

Технологические средства анализа имиджформирующих политических медиатекстов: возможности и ограничения



А. А. Хуажев

Поволжский институт управления имени П. А. Столыпина – филиал РАНХиГС при Президенте РФ, Россия, 410012, г. Саратов, ул. Московская, д. 164

Хуажев Аркадий Аликович, аспирант кафедры политических наук, ooo_zet@mail.ru, <https://orcid.org/0000-0002-1213-9750>

Аннотация. В статье представлены возможности анализа имиджформирующих текстов с использованием специализированного программного обеспечения. Отдельно артикулированы ограничения, связанные с автоматизированной обработкой информации. Делается вывод о предпочтительности применения смешанного варианта исследования медиасообщений, предполагающего сочетание ручного и машинного анализа их основных параметров.

Ключевые слова: политический имидж, имиджформирующий текст, программное обеспечение, интеллектуальный анализ текстов, автоматизированный анализ

Для цитирования: Хуажев А. А. Технологические средства анализа имиджформирующих политических медиатекстов: возможности и ограничения // Известия Саратовского университета. Новая серия. Серия: Социология. Политология. 2022. Т. 22, вып. 2. С. 232–236. <https://doi.org/10.18500/1818-9601-2022-22-2-232-236>

Статья опубликована на условиях лицензии Creative Commons Attribution 4.0 International (CC-BY 4.0)

Article

Technological means of analyzing image-forming political media texts: Possibilities and limitations

A. A. Khuazhev

Stolypin Volga Region Institute of Administration of the Russian Presidential Academy of National Economy and Public Administration, 164 Moskovskaya St., Saratov 410012, Russia

Arkady A. Khuazhev, ooo_zet@mail.ru, <https://orcid.org/0000-0002-1213-9750>

Abstract. The article presents possibilities of analyzing image-forming texts by means of specialized software. Restrictions related to automated information processing are separately articulated. The author concludes that it is preferable to use a mixed version of the study of media messages, involving a combination of manual and machine-based analysis of their main parameters.

Keywords: political image, image-forming text, software, text mining, computer-aided analysis

For citation: Khuazhev A. A. Technological means of analyzing image-forming political media texts: Possibilities and limitations. *Izvestiya of Saratov University. Sociology. Politology*, 2022, vol. 22, iss. 2, pp. 232–236 (in Russian). <https://doi.org/10.18500/1818-9601-2022-22-2-232-236>

This is an open access distributed under the terms of Creative Commons Attribution 4.0 International License (CC-BY 4.0)

В современном информационно насыщенном мире политический имидж во многом формируется посредством медийного сопровождения деятельности различных субъектов и акторов или же за счет целенаправленно создаваемых для этого текстов¹. Подобного рода тексты принято называть имиджформирующими. А. Е. Татарова и Л. Е. Ильина определяют их как «законченные, целостные речевые произ-

ведения, содержащие информацию, направленную на формирование имиджа объекта»².

Несмотря на достаточно устойчивый интерес российских ученых к самым различным аспектам формирования имиджа³, вопрос, связанный с технологической стороной анализа имиджформирующих текстов, нам представляется недостаточно изученным. По нашим наблюдениям, в большинстве случаев для этого ис-



пользуются традиционные качественные методы анализа, которые при всей их продуктивности несут в себе элемент субъективизма. В конечном итоге это может привести к снижению валидности и достоверности полученных с их помощью научных результатов. В этом смысле цель данной статьи мы видим в том, чтобы выявить исследовательские возможности автоматизированных средств анализа текстов и обозначить связанные с их применением ограничения.

В той или иной степени все варианты автоматической обработки медиасообщений являются разновидностями так называемого интеллектуального анализа текста. Данный термин является контекстным переводом на русский язык англоязычного понятия «text mining», которое дословно может быть переведено как «разработка (подобно добыче полезных ископаемых) текста».

В самом общем виде интеллектуальный анализ текста определяют как направление в сфере искусственного интеллекта, целью которого является «получение информации из неструктурированных текстовых данных путем их преобразования в пригодный для дальнейшей работы набор структурированных данных, представленных в удобном виде для машинной обработки»⁴. В настоящий момент интеллектуальный анализ текста достаточно часто используется в бизнес-сфере, государственном управлении, деятельности, связанной с разведкой и обеспечением безопасности. Весьма значительным представляется потенциал данной технологии и применительно к анализу имиджформирующих медиасообщений в научных целях.

Мы провели анализ базовых параметров наиболее известных в России систем интеллектуального анализа текста. Ниже будут представлены основные результаты такого мониторинга: функционал конкретной системы, стоимость ее использования и иные особенности, в первую очередь связанные с тем, на какую аудиторию она преимущественно ориентирована.

Так, одной из самых авторитетных и популярных в нашей стране является система мониторинга и анализа СМИ и социальных медиа «Медialogия»⁵. Приоритетным вектором в совершенствовании методов интеллектуального текстового анализа в компании считают оценку тональности сообщений. Созданная для этого модель использует технологии глубокого обучения – нейронные сети, которые могут оценивать не только сами тексты, но и характеристики упоминаемых в них персон, объектов и организаций.

В качестве отдельного параметра анализа может быть использован механизм оценки яркости или заметности публикации в информационном поле. Что же касается стоимости ис-

пользования возможностей «Медialogии», то она рассчитывается индивидуально, по запросу – в зависимости от объема и параметров отдельного исследования. Однако какими бы ни были расценки, компания, к сожалению, в принципе не работает с физическими лицами. Такой ответ был получен нами в ходе электронной переписки с менеджерами «Медialogии». В основном ее клиентами являются пресс-службы и PR-отделы коммерческих структур или органов власти.

Еще одна компания – это система мониторинга и анализа социальных медиа и СМИ «Brand Analytics». Разработанная ее специалистами система позволяет аккумулировать данные «с наилучшей полнотой и качеством со всех источников – соцсетей, видеохостингов, сервисов карт, публичных каналов и чатов мессенджеров, отзовиков, форумов, блогов, маркетплейсов, онлайн-СМИ, газет, журналов и федеральных телеканалов»⁶. «Brand Analytics» дает возможность автоматически определять тональность упоминаний интересующего заказчика объекта с точностью до 85–90%. Используемые в системе алгоритмы машинного обучения категоризируют сообщения, выявляют тренды медиополя, агрессию, распознают тексты на изображениях и др.

Стоимость услуг компании варьируется от 14 000 до 110 000 руб. в месяц. Чем выше цена, тем больше возможностей появляется у пользователя. Прежде всего, это касается количества анализируемых сообщений в месяц (от 5000 до 350 000), тем (от 3 до 15), тегов в теме (от 30 до неограниченного числа), наличия или отсутствия доступа к материалам офлайн-СМИ и закрытым лентам информационных агентств, возможности осуществлять поиск по архиву за год и выстраивать рейтинги персон и компаний по каждой теме.

Технологии искусственного интеллекта использует также платформа для аналитики социальных медиа «YouScan». Ее отличительной чертой является то, что она располагает «лучшими в индустрии возможностями анализа изображений»⁷. Кроме этого, компания в большей мере, чем конкуренты, ориентирована на обслуживание бизнес-структур. Поиск упоминаний организаций, продуктов или даже маркетинговых мероприятий осуществляется по словам, хэштегам, геолокации и даже по фотографиям. Как и большинство аналогичных сервисов, «YouScan» дает возможность определить тональность медиасообщений. На этот раз декларируемая точность подобной операции составляет 95%.

На сайте компании нет точной информации о расценках на оказываемые ею услуги. Общение с сотрудниками службы поддержки клиентов дало возможность выяснить, что стоимость исследования зависит от многих факторов: количе-



ства объектов мониторинга (брендов, продуктов, персон, кампаний), объема данных по каждому из них, глубины ретроспективы и подхода к обработке упоминаний. В ноябре 2021 г. стоимость проекта, направленного на анализ имиджформирующих текстов в отношении одного объекта, начиналась от 35 000 руб.

Преимущественно на бизнес-структуры ориентирован и ряд других технологических систем, например «SemanticForce». Она представляет собой платформу для мониторинга и анализа интернет-СМИ, социальных сетей, видео, форумов, блогов, вакансий и других видов онлайн-медиа⁸. Используемые «SemanticForce» метрики позволяют оценить частоту упоминаний в социальных медиа бренда, фирмы или отдельного человека за интересующий исследователя временной период. При этом упоминания могут встречаться в твитах, постах, обновлениях статусов, обсуждениях или комментариях. Имеются опция автоматической рубрикации найденных упоминаний и возможность добавления новых рубрик.

Как и большинство других подобных сервисов, «Semantic Force» определяет тональность, однако в данном случае она фиксируется не для всего медиасообщения, а применительно к конкретному объекту, что создает техническую возможность формирования выборок с различной тональностью. Кроме того, если в подавляющем большинстве других сервисов используются три разновидности тональности (положительная, нейтральная и отрицательная), то здесь появляется возможность фиксации параметров, отражающих эмоциональные состояния авторов постов в социальных медиа: счастье, любовь, печаль и гнев.

Информации о стоимости услуг «Semantic Force» на сайте нет. Судя по всему, в каждом отдельном случае она рассчитывается индивидуально, и с учетом ярко выраженной ориентации этой компании на бизнес-структуры можно предположить, что доступ к возможностям системы будет никак не дешевле, чем в сервисах, рассмотренных выше.

Анализ и управление репутацией клиента в Интернете считает своей основной задачей система мониторинга соцмедиа «IQBuzz». Позиционируя себя в качестве «идеального инструмента в SMM-маркетинге», компания занимается оптимизацией деятельности структурных подразделений, отвечающих за маркетинг и PR⁹. Среди прочего в режиме реального времени отслеживаются все упоминания о компании в интернет-среде, имеется возможность определения тональности сообщения – позитивной, нейтральной, негативной и смешанной.

Стоимость услуг «IQBuzz» варьируется в зависимости от тарифа: в январе 2022 г. она составляла 9480 руб. в месяц за «Старт», 17 880 – за «Стандарт», 33 480 – за «Профи» и 65 880 руб. – за «Корпоративный». Набор услуг, включенных в конкретный пакет, в целом аналогичен тому, что упоминался нами применительно к «Brand Analytics». Компания также предоставляет семидневный бесплатный демо, однако без доступа к онлайн-СМИ.

На работе с репутацией клиента в интернет-пространстве фокусируется и система мониторинга упоминаний в социальных медиа «Vabkee». Своей приоритетной целевой аудиторией сервис считает тех, «кто постоянно следит за изменениями, происходящими в определенной сфере бизнеса; потребностями аудитории и нуждается в своевременном проведении анализа рынка конкурентов»¹⁰. Система дает возможность осуществлять анализ эффективности рекламных кампаний в Сети и оказывать оперативную поддержку пользователям.

«Vabkee» – это новый продукт молодой и инновационной команды Ingate Development. В его проектировании и создании приняли участие приглашенные консультанты и эксперты крупных digital-компаний. Однако, несмотря на все это, интерфейс сайта «Vabkee» мы нашли, пожалуй, наименее удобным и функциональным из всех, с которыми приходилось знакомиться. Информации по платным тарифам найти не удалось. Предусмотрена возможность бесплатного месячного доступа к возможностям системы с ограниченным количеством сообщений по теме. Решив воспользоваться такой возможностью, мы, во-первых, не обнаружили функции фильтра источников, по которым нужно было осуществлять поиск интересующих нас материалов, и, во-вторых, простейший поисковый запрос по теме «Северный Кавказ» не дал ни одного результата...

По нашим наблюдениям, в последнее время на российском рынке все более востребованными становятся предложения аналогичных сервисов от зарубежных разработчиков. Один из них – это сервис MAXQDA, адаптированный под особенности русского языка. Он представляет собой профессиональное программное обеспечение для качественных и смешанных методов анализа данных¹¹. MAXQDA совместим с операционными системами Windows и Mac, что делает возможным его применение практически в любой части мира. Технология позволяет анализировать скрипты фокус-групп, интервью и докладов, таблицы, онлайн-опросы, видео- и аудиофайлы, тексты и многое другое.



Существенным отличием функционала MAXQDA от рассмотренных выше программ является то, что пользователь вынужден самостоятельно генерировать коды, находить и загружать в систему корпус текстов, а также кодировать каждый документ. По сути, программа лишь немного упрощает обработку результатов – например, выводит все одинаковые коды из всех текстов в один документ и автоматически формирует облака слов, диаграммы, графики и т.д. Стоимость использования полной версии программы начинается с 47 долларов за полгода.

Таковы наиболее часто встречающиеся на отечественном рынке программные продукты, позволяющие проводить интеллектуальный анализ текста. Что их объединяет?

Прежде всего, это возможность автоматического поиска текстового контента по заданным пользователем тематическим параметрам. Такой функционал есть у всех из рассмотренных выше сервисов, кроме MAXQDA. Данная функция, безусловно, серьезно облегчает работу исследователя и позволяет собрать максимально полный массив документов, в которых упоминается конкретный объект, персона или общность. Причем в область поиска в большинстве случаев попадают как онлайн-СМИ, так и социальные сети, блоги, сайты отзывов и т.д.

Кроме того, технологии автоматически обрабатывают найденный материал (нивелируя тем самым влияние субъективного фактора) и представляют его в презентабельной форме – в виде графиков, диаграмм, облаков слов и т.д.

Вместе с тем, наряду с очевидными преимуществами подобных сервисов, нужно отметить и ряд ограничений, с которыми ученому придется сталкиваться при их использовании. Прежде всего, это относительно высокая стоимость таких продуктов. Представляется, что применять их на постоянной основе (или же хотя бы в течение более или менее продолжительного времени, за которое реально провести масштабное исследование) могут себе позволить лишь достаточно крупные государственные или бизнес-структуры. Исследователям-физическим лицам для приобретения соответствующего программного обеспечения, скорее всего, потребуются грантовая поддержка.

Помимо этого, в большинстве случаев технологии ориентированы на потребности коммерческих структур по анализу репутации, сильных и слабых сторон конкретного бренда, мониторингу отзывов потребителей, сравнению характера представленности собственной продукции с предложениями конкурентов и т.д. Что же касается сугубо политического анализа, то в подавляющем большинстве случаев соответствующий

инструментарий развит слабо. Приходится либо довольствоваться тем, что есть, либо просить разработчиков адаптировать имеющийся функционал под нужды конкретного исследования (за отдельную плату, разумеется).

Опыт использования демоверсий подобных систем свидетельствует о том, что иногда у пользователя может не быть доступа к самим исходным данным. Иными словами, ученый получает только финальные отчеты, но при этом не может посмотреть конкретные тексты, на основе которых программа сгенерировала какие-либо выкладки и количественные данные. На наш взгляд, это очень существенно снижает возможности верификации получаемых таким образом результатов.

Плюс ко всему, очевидно, что любые автоматизированные алгоритмы, как правило, очень успешно обрабатывают числовую информацию, но при этом не всегда столь же эффективны при их качественной, смысловой интерпретации. В этом смысле мы склонны согласиться с С. В. Ямщиковым и Э. С. Ивановым, указавшими на необходимость «антропологической коррекции полученных машинным способом данных»¹². С другой стороны, сам факт того, что данные, которые выдает машина, в конечном счете все равно анализирует и оценивает сам ученый, говорит о том, что присущий «ручному» способу обработки текстов человеческий фактор, по большому счету, присутствует и здесь (хотя и в чуть меньшем масштабе).

Наконец, нужно обратить внимание и на еще одну потенциальную уязвимость автоматизированного анализа данных. Имеем в виду возможные ошибки и несовершенства, заложенные в само программное обеспечение. Не имея доступа к поисковым алгоритмам, ученому приходится полагаться на то, что они составлены корректно, и система, условно говоря, находит все упоминания интересующего его объекта, верно определяет тональность сообщений, правильно считает и т.д. А что, если на уровне программирования были допущены какие-либо ошибки?

Принимая все это во внимание, мы полагаем, что наиболее оптимальным является использование смешанного или полуавтоматизированного режима анализа. Он предполагает, что сбор и обработка имиджформирующих текстов ведется вручную, однако при этом программное обеспечение тоже используется, но в ограниченном объеме и так, чтобы была возможность самому проверить корректность автоматически генерируемых данных. Тем более что существует целый ряд сервисов, которые дают ученому возможность измерить отдельные характеристики медиасообщений, но при этом их использование



совершенно бесплатно (разработчики получают деньги от рекламы, размещенной на таких сайтах). При этом нужно отметить тот факт, что возможности подобных алгоритмов уже были апробированы отечественными учеными¹³.

Разумеется, в таком варианте процесс анализа становится гораздо более трудоемким, но зато, как нам представляется, и более надежным. Вручную получится обработать меньшее количество медиасообщений, однако глубина такого анализа должна быть явно выше, чем в случае с автоматизированной обработкой текстов. В этом плане мы полностью разделяем позицию А. И. Мицкевича, полагающего, что даже небольшое количество источников при их правильном подборе может адекватно отразить общую ситуацию и привести к конкретным выводам¹⁴.

Что именно предполагает смешанный формат обработки имиджформирующих текстов? На наш взгляд, в самом общем виде он может включать в себя следующие этапы:

- выбор релевантных исследовательским задачам источников информации;
- поиск по ключевым словам сообщений по конкретной тематике;
- фиксация элементов медийных повесток дня первого (о чем сообщают в целом) и второго (что именно пишут или говорят по конкретной теме) уровней;
- отслеживание динамики появления медийных материалов по конкретной тематике;
- выявление количественных параметров медиатекстов: уровня их удобочитаемости, «водности», «тошноты» и степени соответствия закону Ципфа;
- определение тональности текстов в отношении интересующего ученого объекта и на этой основе измерение степени агрессивности информационной среды.

Убеждены, что осуществленное подобным образом исследование позволит вскрыть ключевые характеристики имиджформирующих текстов, что, в свою очередь, даст ученому возможность приблизиться к пониманию механизмов их влияния на массовое политическое сознание аудитории.

Примечания

¹ См., например: *Вилков А. А., Тимофеев Е. И.* Политическая технология брендинга региона как возможность формирования его положительного имиджа и повышения инвестиционной привлекательности // *Известия Саратовского университета. Новая*

серия. Серия : Социология. Политология. 2019. Т. 19, вып. 3. С. 304–311. <https://doi.org/10.18500/1818-9601-2019-19-3-304-311>; *Вилков А. А.* Особенности имиджа России в условиях обострения международных отношений // *Международный и национальный механизмы обеспечения суверенитета : материалы VIII Междунар. Констиуц. Форума, посвященного 80-летию Саратовской области* (Саратов, 16 декабря 2016 г.). Саратов : Саратовский источник, 2017. С. 53–56.

² *Татарова А. Е., Ильина Л. Е.* Имиджформирующий текст как психолингвистическое явление // *Вестник современных исследований*. 2018. № 8.1 (23). С. 372.

³ См., например: *Важенина И. С., Важенин С. Г.* Имидж как конкурентный ресурс региона // *Регион : экономика и социология*. 2006. № 4. С. 72–84; *Напалкова И. Г., Курочкина К. В.* Имидж Республики Мордовия : образы, символы, стереотипы // *Регионология*. 2020. Т. 28, № 4 (113). С. 866–887. <https://doi.org/10.15507/2413-1407.113.028.202004.866-887>; *Панина Е. А., Дунаевская Е. А.* Брендинг как технология создания имиджа региона (на примере Республики Адыгея) // *Вестник Майкопского государственного технологического университета*. 2012. № 3. С. 125–128.

⁴ *Долгачева Е. Л., Косюк Е. Ю., Попова Д. Л., Русаков А. М.* Подбор инструментальных средств анализа текстов для создания информационно-аналитической системы // *Научный альманах*. 2020. № 11-2 (73). С. 28.

⁵ См.: *Медиалогия : О компании*. URL: <https://www.mlg.ru/about/> (дата обращения: 10.01.2022).

⁶ *О Brand Analytics*. URL: <https://br-analytics.ru/about/> (дата обращения: 10.01.2022).

⁷ *YouScan : О продукте*. URL: <https://youscan.io/ru/> (дата обращения: 10.01.2022).

⁸ См.: *SemanticForce : About us*. URL: https://semanticforce.net/ru/about/about_us/ (дата обращения: 11.01.2022).

⁹ См.: *IQBuzz : О сервисе*. URL: <https://iqbuzz.pro/about.php> (дата обращения: 11.01.2022).

¹⁰ *Babkee : О нас*. URL: <http://www.babkee.ru/about/> (дата обращения: 11.01.2022).

¹¹ См.: *О MAXQDA*. URL: <https://www.maxqda.com/russia/kachestvennogo-analiza-dannykh> (дата обращения: 11.01.2022).

¹² *Ямицков С. В., Иванов Э. С.* Системы мониторинга интернет-медиа как инструмент социологического исследования виртуальных социальных сетей // *Социосфера*. 2021. № 1. С. 105.

¹³ См., например: *Казаков А. А.* Политическая теория и практика медийной грамотности. Саратов : Изд-во Саратовского ун-та, 2019. 172 с.; *Его же.* Результаты апробации методики измерения манипулятивного потенциала политических медиатекстов // *Информационные войны*. 2020. № 3 (55). С. 89–94.

¹⁴ См.: *Мицкевич А. И.* Политическая медиаметрия : управленческо-прикладные аспекты // *Научные труды Республиканского института высшей школы*. 2019. № 18. С. 204.

Поступила в редакцию 15.01.2022; одобрена после рецензирования 25.01.2022; принята к публикации 05.02.2022
The article was submitted 15.01.2022; approved after reviewing 25.01.2022; accepted for publication 05.02.2022