

Богомолова Татьяна Юрьевна
Институт экономики и организации
промышленного производства СО РАН,
Новосибирский национальный исследовательский
государственный университет,
г.Новосибирск, Российская Федерация
bogtan@rambler.ru

Щербакова Дарья Михайловна
Новосибирский национальный исследовательский
государственный университет,
г.Новосибирск, Российская Федерация
dasha_sh54@icloud.com

Структурирование и анализ больших данных: опыт изучения проблемного поля онлайн-петиций на портале change.org

Аннотация. Данное исследование представляет собой попытку рассмотрения как «электронной демократии», так и её конкретного механизма «электронной петиции» с целью дополнения или расширения существующего социологического знания о способах и формах коммуникации между населением и властными структурами в цифровую эпоху. В исследовании используются новые доступные для современной науки методы работы с большими данными, а также традиционные техники преобразования неструктурированных данных в структурированные для применения традиционных способов статистической обработки.

Ключевые слова: электронная демократия; онлайн-петиции; Большие данные; методы анализа эмпирических данных

Bogomolova Tatyana Yuryevna
Institute of Economics and
Industrial Engineering SB RAS,
Novosibirsk National Research State University,
Novosibirsk, Russian Federation
bogtan@rambler.ru

Shcherbakova Daria Mikhailovna
Novosibirsk National Research State University,
Novosibirsk, Russian Federation
dasha_sh54@icloud.com

Structuring and analysis of big data: experience of studying the problem field of online petitions on the change.org portal

Abstract. This study is an attempt to consider both “electronic democracy” and its specific mechanism of “electronic petition” in order to supplement or expand existing sociological knowledge about the ways and forms of communication between the population and government in the digital age. The study uses new methods of working with big data available

for modern science, as well as traditional techniques for converting unstructured data into structured ones for applying traditional methods of statistical processing.

Keywords: e-democracy; online petitions; Big data; empirical data analysis methods

В последнее время все более популярными среди граждан разных стран становятся площадки онлайн-петиций, используемые с целью решения проблем населения путем формирования общественной поддержки. Как утверждается в аналитическом отчете портала петиций change.org, «каждый месяц создается более 25 000 петиций, затрагивающих самые разные проблемные сферы», а число пользователей к 2018 году достигло цифры в 265 786 771 пользователей по всему миру [Change.org., 4]. В нашей стране, как указывается, число зарегистрированных пользователей также растет и в настоящее время составляет 14,3 млн пользователей [Change.org, 4]. Вместе с этим увеличивается и разнообразие существующих платформ, предлагающих такую возможность: «РОИ», «Change.org», «Online Petition.ru», «Демократор.ру», «Alter-Russian Federation» и другие.

Данное исследование представляет собой попытку рассмотрения как «электронной демократии», так и её конкретного механизма «электронной петиции» с целью дополнения или расширения существующего социологического знания, используя при этом новые доступные для современной науки методы работы с большими данными. Предполагается, что полученные выводы в ходе данного исследования позволят сформировать представление о результатах внедрения элементов «электронной демократии», которые нацелены на решение проблем общества, выносящихся на публичное обсуждение и требующих участия властных органов.

Цель работы – выявить проблематику и характеристики онлайн-петиций населения, в том числе тех, требования которых удовлетворены властными структурами («победившие» петиции).

Эмпирическим объектом исследования являются онлайн-петиции на портале change.org. Выбор данного портала обусловлен целым рядом его преимуществ в сравнении с другими российскими аналогичными порталами. Здесь отметим лишь одно ключевое преимущество: онлайн-петиции, размещенные на change.org имеют большую степень открытости по сравнению с петициями других порталов. Это выражается в том, что можно понять, кто является инициатором петиции, кто её поддержал, кто ею поделился и кому она обращена/направлена.

В исследовании анализировались петиции, размещенные с начала времени функционирования данного портала на просторах российского Интернета до наших дней – с 2013 по 2019 год. Для решения одних задач исследования данные портала использовались в исходном неструктурированном виде (Табл. 1). В этом случае к ним применялась техника обработки естественного языка (Natural Language Processing) с фокусированием на методе тематического моделирования (среда Spyder, язык программирования – Python). Для решения других задач были сформированы

несколько выборок из генеральной совокупности петиций. Каждая петиция из выборочной совокупности выступила единицей анализа и была описана следующим перечнем индикаторов: характеристики проблемы, поднятой в петиции; характеристики субъектов, к которым обращена петиция; характеристики инициаторов петиции; характеристики текста петиций; характеристики официальных ответов на петиции; решение проблемы, поднятой в петиции; масштаб активности населения, направленной на поддержку решения данной проблемы посредством подписания онлайн-петиции. В итоге все выборочные массивы были представлены в традиционном для стандартных статпакетов формате «объект-признак».

Таблица 1. Характеристика информационной базы исследования

Вид данных	Среда представления	Объем массива, петиции, шт.
Массив всех петиций (генеральная совокупность, неструктурированные текстовые данные)	Среда Spyder (язык программирования – Python)	Свыше 300 000
Массив «победивших» петиций (генеральная совокупность, неструктурированные текстовые данные)	Среда Spyder (язык программирования – Python)	88
Выборочный массив из всех петиций, за исключением «победивших» (неструктурированные текстовые данные)	Среда Spyder (язык программирования – Python)	100
Выборочный массив из всех петиций, включая «победившие» (структурированные данные, представленные в формате «объект-признак»)	SPSS	100
Выборочный массив из всех петиций за исключением «победивших» (структурированные данные, представленные в формате «объект-признак»)	SPSS	100
Массив «победивших» петиций (генеральная совокупность, структурированные данные, представленные в формате «объект-признак»)	SPSS	88

Здесь представим пример того, как соотносятся между собой результаты исследования одной и той же совокупности объектов с использованием разных аналитических техник. Применение такого метода анализа текстов как тематическое моделирование к полной совокупности петиций, размещенных в 2013–2019 гг. на портале change.org, позволяет выделить четыре блока поднимаемых в онлайн-петициях тем: проблемы защиты и охраны животных; вопросы здравоохранения и оказания медицинской помощи, особенно детям и инвалидам; проблемы соблюдения законов и реализации прав граждан; проблемы, связанные с территориальным развитием. Анализ выборочной совокупности петиций показал, что наиболее часто в онлайн-петициях поднимаются проблемы защиты и охраны животных, а также проблемы, связанные с вопросами здравоохранения и оказания медицинской помощи, а, значит, чаще всего петиции адресуют в Министерство здравоохранения, Министерство природы и Прокуратуру. Победившие петиции, т. е. те петиции, требования которых были

удовлетворены, поднимают преимущественно одну проблему: вопросы здравоохранения и оказания медицинской помощи детям, и ситуации взаимодействия, направленной на эту помощь: ребенок – родитель – врач.

При триангуляции различных методов анализа больших данных появляется шанс многомерного описания предмета исследования – в нашем случае онлайн-петиций как формы коммуникации в рамках института обращений в цифровую эпоху.

Библиографический список

Change.org в мире: победы и достижения 2018 [Электронный ресурс] // URL: https://static.change.org/brandpages/impact/reports/2019/change.org_Impact_Report_russian_FINAL.pdf (дата обращения: 19.02.2020).