

И. И. Зубова

РЕПУТАЦИЯ ЛИЧНОСТИ И ЕЕ АВТОМАТИЧЕСКАЯ ИДЕНТИФИКАЦИЯ

В статье рассмотрены основные составляющие репутации политического деятеля и особенности ее отражения в субъективных высказываниях людей. Представлена база данных (набор графических, грамматических, лексических, визуальных средств) и формальная модель (алгоритм) системы автоматического анализа репутации личности Президента США Д. Трампа на основе мнений, выраженных в текстах англоязычных твитов. В ходе проверки адекватности работы формальной модели сформулирован вывод о том, что в целом она разработана правильно, однако ее можно улучшить путем расширения базы данных и использования ряда дополнительных лексических и синтаксических правил анализа единиц твита.

Ключевые слова: алгоритм, база данных, личность, маркер, репутация, твит.

I. Zubova

PERSONAL REPUTATION AND ITS AUTOMATIC IDENTIFICATION

The article considers the main components of a politician's reputation and the specific features of its reflection in the subjective statements of people. The article describes a database (a set of graphical, grammatical, lexical, and visual means) and a formal model (algorithm) of the automatic analysis of D. Trump's reputation. The model is based on the opinions about the US President expressed in English tweets. The formal model adequacy test showed that the model is, in general, effective. However, it may benefit from a bigger database and additional lexical and syntactic rules for the analysis of tweets.

Keywords: algorithm, database, personality, marker, reputation, tweet.

Автоматическое определение субъективности (оценочности/тональности) интернет-текста является одним из динамично развивающихся направлений современной компьютерной лингвистики, поскольку оперативный и объективный анализ субъективного образа исследуемого объекта или субъекта, возникающего в ходе естественных процессов коммуникации или намеренно создаваемого в информационном поле сети Интернет, является важной и востребованной частью проведения политических кампаний, обеспечения безопасности общества в целом и его субъектов в частности. Это также значимая составляющая определения эффективности рекламных кампаний, PR-кампаний и прочих видов манипуляций современным общественным сознанием. Развитие политтехнологий и технологий информационной раз-

ведки в современном обществе вызывает известный интерес к проблеме определения оценочности текста, а сквозь призму оценочности — к проблеме формирования репутации одного человека, группы людей или предприятия. Репутация имеет большое значение во всех сферах деятельности, о чем свидетельствует, например, тот факт, что во многих организациях есть PR-отделы, сотрудники которых занимаются вопросами создания и поддержки положительной репутации своих руководителей и предприятия в целом.

Репутация личности и ее составляющие

В повседневной жизни человек иногда сталкивается с необходимостью выбора какой-то организации в качестве места работы, предприятия, предоставляющего определенные услуги, претендента на вакантное место

работы и т. д. Делая этот выбор, он часто опирается на репутацию конкретной личности или организации, т. е. доверяет информации, являющейся результатом выражения мнений группы людей. В «Толковом словаре живого великорусского языка» В. И. Даля репутация определяется как «слава человека, добрая и дурная, как и чем кто слышет, общее мнение о ком-либо или чем-либо» [2, с. 522]. Д. Н. Ушаков понимает под репутацией «состоявшееся общественное мнение о достоинствах или недостатках кого-чего-нибудь» [14, с. 597]. В словаре С. И. Ожегова приводится следующее определение данного понятия: «Репутация — приобретаемая кем-нибудь или чем-нибудь общественная оценка, общее мнение о качествах, достоинствах и недостатках кого-нибудь или чего-нибудь» [11, с. 677]. На основе приведенных выше и целого ряда других определений можно выделить некоторые характеристики репутации.

Прежде всего, репутация не заложена в субъекте или объекте изначально, а приобретается в результате деятельности человека, группы лиц или организации. Кроме того, репутация субъективна, так как формируется на основе мнений или оценок группы людей. Она может быть как положительной, так и отрицательной, относительно устойчива, хотя довольно легко добиться ее изменения в негативную сторону и довольно трудно восстановить некогда положительную, но в результате каких-либо факторов «испорченную» репутацию. Наконец, репутация является одним из наиболее эффективных механизмов социального контроля в обществе. Люди формируют свое мнение, оценивая на основе определенных критериев поведение, качества или характеристики субъекта либо объекта, и затем используют результат оценки для предположения дальнейшего поведения того, мнение о ком/чем они формируют. Базируясь на этом предположении, люди могут скорректировать свое отношение к данному человеку, группе лиц или организации, тем самым решая, следует ли ему/им доверять. Очень часто в качестве синонима понятия «репутация» используется понятие

«имидж». Несмотря на то, что эти понятия довольно близки, между ними есть некоторые различия. В отличие от репутации, имидж формируется искусственным путем и целенаправленно. Это не естественная общественная оценка, а специально созданный с помощью рекламы, пропаганды или PR-кампаний образ субъекта либо объекта. Имидж формируется в течение относительно небольшого промежутка времени, легко изменяется, довольно поверхностен и базируется скорее на впечатлениях о внешних характеристиках субъекта или объекта, в то время как репутация подразумевает его более глубокую и осмысленную оценку [3].

Выделяют два уровня репутации. Во-первых, следует отметить индивидуальный уровень, т. е. репутацию личности (конкретного индивида) как «субъекта деятельности, в единстве его индивидуальных свойств и социальных ролей» [5, с. 7]. В более узком смысле под персональной репутацией понимают «тот спектр характеристик человека, который учитывает все факторы внешнего воздействия и существующие тенденции, имеющие отношение к его деятельности, личным качествам, а также когнитивной составляющей, и который формируется на основе общественного мнения и общепринятых стандартов» [15]. Таким образом, репутация личности рассматривается как «феномен публично отфильтрованного мнения, оценка нравственных, деловых, профессиональных качеств личности» [13, с. 10]. Персональная репутация важна в любой сфере деятельности, поскольку дает человеку возможность работать там, где его профессионализм подтвержден репутацией, и при этом с профессиональной точки зрения занимать выгодные позиции. Так, в политической сфере репутация стала выделяться как одна из важных составляющих политического успеха или провала политического лидера либо партии (например, в ходе выборов при равных условиях побеждает политик с преобладанием позитивных характеристик в глазах общественности). Во-вторых, необходимо отметить надиндивидуальный уровень, т. е.

репутацию группы лиц, коллектива, сообщества, организации, страны, нации.

Персональная репутация политического лидера, как выраженная в общественном мнении система качеств обладающей властью личности, складывается из определенных структурных компонентов: нравственного, интеллектуального, образовательного, профессионального, психологического. Большое влияние на формирование репутации политика оказывает генеалогия, семья (в том числе супруг или супруга), уровень его материального обеспечения (размер доходов), дружественные и деловые связи и т. д. [13]. А. Ю. Трубецкой отмечает, что репутация политического лидера имеет устойчивую структуру, которая включает в себя следующие основные психосемантические факторы: 1) предназначение власти — описывает отношение политического лидера к своей стране, государству, к своему народу; 2) силу личности — включает в себя волевую регуляцию поведения, решительность, принципиальность, трудолюбие и интеллект; 3) нравственно-этическую позицию лидера — описывает моральные и душевные качества и ценности политика. Дополнительными факторами являются выразительность самопрезентации, стиль руководства, коммуникативная установка, религиозность, тип мышления и мускулиность — феминность [12, с. 23–25]. С понятием репутации тесно связаны такие явления, как уважение — почтительное отношение, основанное на признании чьих-либо достоинств, вид одобрения поведения и качеств лица, занимающего определенный социальный статус и имеющего социальную значимость; реноме — установившаяся положительная репутация; авторитет — общепризнанное влияние, вес, значение, которым пользуется определенный человек, группа людей, организация вследствие своих достижений, качеств и заслуг; престиж, который определяет место в социальной иерархии, или социальный статус, который имеет конкретное лицо или организация в иерархии других статусов [6].

С лингвистической точки зрения основой формирования репутации субъекта или объ-

екта являются субъективные высказывания, содержащие субъективные мнения или оценки достаточно большой группы людей. Рассмотрим репутацию сквозь призму категории субъективности текста подробнее.

Субъективное высказывание как основа формирования репутации личности

В различных научных областях понятие «субъект» определяется по-своему. Так, в философии он имеет довольно обобщенное понимание. С философской точки зрения «в основе категории субъекта лежат такие понятия, как сущность, индивид, обладающие сознанием и волей, познающие мир и практически преобразующие его» [9]. В психологии под этим термином понимают «индивида или группу как источник познания и преобразования действительности; носителя активности, осуществляющего изменения в других людях и в себе самом, как другом. Субъектность человека проявляется в его витальности, деятельности, общении, самосознании» [9]. С точки зрения морфологии *субъект* — это корень, от которого происходят родственные слова: 1) *субъективность*, т. е. конкретные представления человека обо всем, что его окружает, основанные на его чувствах, мыслях, ощущениях (иначе говоря, это точка зрения, сформированная в результате полученных знаний и размышлений, мировоззрение); 2) *субъективное*, т. е. личное, внутреннее состояние, переживания. Данная категория свидетельствует также о взаимодействии людей между собой и с окружающей действительностью, их иллюзиях и заблуждениях [8]. Как неоднократно отмечалось ранее, репутация является общественным мнением, которое складывается из мнений отдельных людей. В этом контексте возрастает важность каждого отдельно взятого высказывания, в котором присутствует субъективная оценка кого-либо или чего-либо. Многие лингвисты отмечают, что главным отличием текстов, где присутствует субъективная доминанта, от информативных текстов является то, что в понимании их смысла большую роль играет субъективный фактор, тесно связанный

с выражением мнения автора, другими словами, с субъективной модальностью, которая «есть субъективное отношение производителя речи к содержанию высказывания (текста)» [10, с. 61].

Единое общепринятое определение понятия «субъективность» еще не сформулировано. Однако ряд ученых, в том числе и В. Хатзивассилоглу, понимают под субъективностью «всю информацию, которая относится к мнению и оценке» и противопоставляют ей объективную информацию, которая «сообщает факты об объектах действительности» [16, с. 300]. Например, предложение *The plot of the movie “Arrhythmia” unfolds around the ambulance doctor* является объективным, так как не содержит оценочную информацию. С другой стороны, предложение «*Arrhythmia*» *is a very good film! It is a surprisingly lively film, bribing with its simplicity and recognizability of archetypes* содержат оценочную информацию и являются субъективными. Необходимо подчеркнуть, что термин «субъективность» близок к термину «оценка», который употребляется в лингвоаксиологии. В процессе коммуникации человек постоянно оценивает окружающие его реалии, других людей, себя, выражая свое мнение и отношение разнообразными языковыми средствами. Важно отметить, что «субъективный характер оценки заключается в наличии субъекта (индивидуума или коллектива), применяющего общественную или личную систему ценностей в зависимости от собственных потребностей или намерений, которые определяют прагматику его высказываний» [1, с. 66]. Человеческое мышление может быть субъективным и объективным. Объективное мышление полагается на те данные, которые научно обоснованы, экспериментальны, одним словом, объективны. Субъективное мышление, в свою очередь, основано на чувствах, эмоциях, желаниях и оценках. А. Н. Леонтьев определяет эмоции как состояния, выражающие «оценочное личностное отношение к складывающимся или возможным ситуациям, к своей деятель-

ности и своим проявлениям в них» [7, с. 37]. Общеизвестно, что мышление взрослого, здорового человека неразрывно связано с речью, в том числе и с письменной. Невозможно представить, чтобы мысль возникала и существовала вне языка, вне речи. Следовательно, можно говорить о том, что субъективное мышление находит свое отражение в том, что и как человек пишет или говорит. А это отражается на выражении своего мнения. В общепринятом представлении под мнением подразумевается взгляд человека на какую-либо проблему, предмет или явление, его оценка или суждение (от старослав. *мьнити — предполагать, думать*). Согласно новому словарю русского языка Т. Ф. Ефремовой, мнение — это «суждение, выражающее оценку кого-либо или чего-либо, отношение к кому-либо или чему-либо, взгляд на кого-либо или на что-либо» [4]. Исходя из вышесказанного, можно заключить, что субъективность в тексте — это выражение мнения его автора о ком-либо или о чем-либо посредством различных языковых средств.

Бурное развитие сети Интернет привело к формированию новых сетевых коммуникационных моделей. Если раньше были распространены такие формы виртуального общения, как чаты, блоги и форумы, то на сегодняшний день пользователи отдают предпочтение платформам, которые совмещают в себе возможности обмена сообщениями между пользователями один на один в режиме реального времени, открытого публичного обсуждения какой-то темы (выражения мнений) множеством пользователей и публикаций сообщений различного содержания, доступных всем пользователям Сети. Социальная сеть Twitter является одной из самых популярных платформ, которая предоставляет все эти возможности. Структурно-смысловым элементом этой сети является твит — короткое сообщение пользователя на определенную тематику (длиной не более 280 символов). Твиты объединяются в одну тематическую цепочку специальным знаком — хэштегом, что позволяет вычленять из огром-

ного потока информации нужные пользователю сообщения. Чаще всего такие сообщения передаются в печатном виде, поэтому в интернет-коммуникации отсутствует паралингвистическая информация (мимика, жесты, тон и тембр голоса и т. п.). Кроме того, общение в Интернете характеризуется несоблюдением орфографических и пунктуационных норм, присущих письменной речи, и тем самым сближается с устной речью. В устной коммуникации человек выражает свою оценку не только вербальными, но и невербальными средствами. С этой целью в интернет-общении вводятся косвенные средства выражения оценки, чаще всего эмодзи и стикеры, позволяющие пользователям, с одной стороны, выражать свои эмоции, формируя при этом оценки (мнения), а с другой стороны, делать виртуальное общение лаконичным.

Ежедневно на разных интернет-площадках появляется огромное количество мнений (оценок) пользователей Сети, поэтому процесс сбора и обработки вербальных и невербальных средств формирования субъективных высказываний виртуальных коммуникантов очень сложно проводить вручную. Рассмотрим возможный вариант моделирования данного процесса на примере создания прототипа системы автоматического анализа репутации Президента США Д. Трампа в англоязычном сегменте сети Twitter.

База данных компьютерной системы

В качестве эмпирического материала были использованы тексты 500 англоязычных твитов, опубликованных с января 2017 года по апрель 2019 года, в которых пользователи сети оценили деятельность и некоторые личностные качества политического лидера США. Необходимо подчеркнуть, что участники сетевого общения зачастую используют формы выражения мыслей, далекие от традиционных лексических норм. Так, в текстах интернет-коммуникации встречаются аббревиатуры, отражающие, по сути, реакцию пользователя (*LOL* — *laughing out loud*), сокращения (*abt* — *about*), намеренные орфо-

графические ошибки, опечатки и т. д. С лингвистической точки зрения лексическая форма их представления далека от традиционной, принятой в конкретном языке, что делает практически невозможным их автоматический анализ до тех пор, пока они не будут нормализованы. Поэтому предварительный анализ материала исследования позволил выявить в массиве твитов слова, подлежащие лексической нормализации, например, *vey*, *wanna*, *w/o*, *some one*, *tbh*, *4 me*, *grt ag* и др. В полном объеме они составили список ненормализованных лексических единиц и правильных вариантов их написания (около 90 пар слов).

В процессе дальнейшего исследования текстов твитов были выделены слова, обозначающие основной субъект определения репутации (общественного мнения) — *Donald Trump*, аспектные категории (характеристики) этого субъекта — *D. Trump in general*, *D. Trump as the President of the USA*, *D. Trump's administration*, *D. Trump's party* и *D. Trump's supporters*, а также связанные с ними аспектные термины (ключевые слова) — *Donald Trump*, *Trump*, *Donald*, *Donny*, *@realDonaldTrump*, *Donald Trump's*; *president*, *presidency*, *policy*, *policies*, *agenda*, *POTUS*, *@POTUS*, *President Trump*; *administration*, *office*, *White House*, *@WhiteHouse*; *the Republican Party*, *Republican*, *Republicans*, *Repub*, *GOP*, *@GOP*; *supporters*, *supporter*, *trumpsters*. Выявление аспектных категорий и аспектных терминов (опорных слов) позволяет исключить из автоматического анализа те лексические единицы, которые не влияют на оценку выбранной политической личности, т. е. на ее репутацию.

На следующем этапе анализа эмпирического материала были выявлены вербальные и невербальные маркеры отражения репутации Д. Трампа. К ним относятся оценочная (эмоционально-оценочная и эмоционально-экспрессивная) лексика, слова-интенсификаторы, слова-инверторы, хэштеги, стоп-слова, а также графические знаки — эмодзи и эмодзи-интенсификаторы. Список

оценочной лексики (около 750 единиц) состоит из двух частей: перечня положительных и перечня отрицательных оценочных (эмоционально-оценочных и эмоционально-экспрессивных) слов с указанием их оценочного веса. В качестве шкалы оценки были выбраны веса от +2 (сверхположительный вес) до -2 (сверхотрицательный вес). Например:

Replying to @realDonaldTrump

*Good morning President Trump!!! **Hope** (+1) you have an even **better** (+2) day today. **Thank** (+2) you for standing **strong** (+1) for the **Patriots** (+2)! We **appreciate** (+2) all that you do!*

*Trump's too **stupid** (-1) to understand policy. He's an **impotent** (-1) and **ignorant** (-1) president **weakening** (-1) our country.*

Replying to @pawtrixcove @wdunlap and 3 others

*The real truth is that Donald J Trump is a **conman** (-2), a **grifter** (-2), a **racist** (-2), an **imbecile** (-2), a **misogynist** (-2), a **cheat** (-1), and an all around **bad** (-1) person. He surrounded himself with **criminals** (-2) and is the ultimate cause of all that has happened to him. It's **unpatriotic** (-2) to look past these things.*

Исследование отобранного массива твитов англоязычных пользователей показало, что их авторы применяют различные интенсификаторы, т. е. лексические единицы, увеличивающие или уменьшающие вес оценочного слова (*amazingly, extremely, consistently, unexpectedly, systematically, regularly* и др.). Выявленные интенсификаторы (55 лексических единиц) были ранжированы по следующей шкале: высокая (оценочный вес 1.25), средняя (оценочный вес 0.5), низкая степень (оценочный вес 0.25) усиления/уменьшения оценки, что будет учитываться при построении соответствующих правил. Например:

Replying to @realDonaldTrump

*Trump is our National Emergency!! We need to keep an eye on him. He is becoming **increasingly** (1.25) erratic and may take drastic actions in an attempt to distract America.*

Replying to @realDonaldTrump

***Actually** (0.5), the world sees us as a weak country because of the weak president we have!*

Replying to @maddow

*Define embarrassing? **Quite** (0.25) **frankly** (0.5) I think literally everything Trump does is embarrassing.*

Помимо оценочной лексики, в англоязычных твитах были выявлены слова (около 20 лексических единиц), инвертирующие вес оценочного слова на противоположное значение (*appropriate +1 — not appropriate -1; embarrassing -1 — never embarrassing +1* и т. п.). Например:

Replying to @MaxBoot

*Trump was and continues to be a national security threat. He makes us **less** safe every day he remains in office.*

*Let's be honest. When tragedy happens, we all look to see how embarrassing trump's reaction will be. He's merely a perverse entertainer with an undeserved platform. **Never** a leader. **Never** a role model. **Never** the best. Always the absolute worst.*

*Trump's own speeches are laced with debunked lies — more than a thousand false claims since he took office. **None** corrected. **No** apologies. His career is built on lies & conspiracy theories. And the free press is a core pillar of democratic accountability — which is why he attacks it.*

В эмпирическом материале были обнаружены оценочные слова, которые относились не к Д. Трампу, а к его оппонентам. Чтобы определение репутации Президента США в англоязычных твитах было максимально точным, лексические единицы, обозначающие других субъектов и объекты, были выделены в отдельный список так называемых стоп-слов. К ним были отнесены следующие единицы: *Obama, Barak Obama, President Obama, President Barak Obama, Obama's; Hillary, Hillary Clinton; Democratic party, Dems, Democrats; MSM, media, mainstream media, news media.*

Как отмечалось ранее, особенностью коммуникации в социальной сети Twitter является использование в сообщениях пользователей специального знака — хэштега и связанной с ним цепочки символов. В ходе анализа массива англоязычных твитов было выяснено, что некоторые хэштеги могут нести в себе определенную оценку и, таким образом, напрямую влиять на общий оценочный вес высказывания. Поэтому каждому выделенному хэштегу был приписан оценочный вес от +1 до -1. Например, *#AmericaFirst* (+1), *#GodBlessTrump* (+1), *#Diplomacy* (+0.5), *#biggestliarever* (-1), *#ImpeachTrump* (-1), *#OnTheDayTrumpResigns* (-0.5). В полном объеме они составили список хэштегов, включающий около 60 лексических единиц.

Как было отмечено выше, косвенным средством выражения оценок, эмоций и психических состояний участников виртуального общения являются эмоджи. Данные знаки влияют на общую репутацию личности, усиливая или уменьшая ее, поскольку обладают оценочным весом от +1 до -1. Например, ❤️ (+1), 💎 (+0.5), 🇺🇸 (+0.25), 😡 (-1), 🙄 (-0.5). Список подобных идеографических знаков включил 25 единиц. Следует отметить, что некоторые эмоджи, хотя и не выражают эмоцию (оценку) напрямую, в то же время могут увеличивать или уменьшать общий вес представленной в тексте твита репутации. Назовем подобного рода графические знаки эмоджи-интенсификаторами. Например, !! (1), 🇺🇸 (0.5), 🙋 (0.5), 👉 (0.25).

Рассмотренные выше языковые и графические маркеры составили базу данных системы автоматического анализа репутации личности Президента США Д. Трампа в англоязычных твитах.

Формальная модель компьютерной системы

На основе сформированной базы данных была разработана формальная модель (алгоритм) прототипа системы. Отметим основные особенности данного алгоритма. Анализ персональной репутации проводится в четыре

этапа. Первый этап заключается в первоначальной обработке текста твита. При этом текст обрабатывается целиком, без разбиения на отдельные предложения. Сначала проводится нормализация его лексических единиц. Далее текст твита анализируется на наличие в нем невербальных знаков эмоджи. При их обнаружении они сопоставляются с элементами графической части базы данных. При совпадении с зафиксированными данными их вес (f_e) добавляется в соответствующий счетчик оценки текста (K_n — для отрицательных оценочных элементов; K_p — для положительных оценочных элементов). При совпадении найденного эмоджи с одним из эмоджи-интенсификаторов его оценочный вес заносится во временную память (f_{ei}). Далее проводится проверка текста на наличие в нем сочетаний знаков препинания. Так как сочетание типа !!!!! или ??? являются косвенным средством выражения оценки, то при их наличии в тексте твита во временную память заносится значение $f_{mi} = 2$. Если в конце предложения стоит один восклицательный знак, то во временную память заносится значение $f_{mi} = 0.5$. На третьем этапе анализа репутации после окончательного вывода о направлении оценки всего твита значения, находящиеся во временной памяти, добавляются в счетчик K_n или счетчик K_p .

На втором этапе проводится основной анализ оценочности твита. Текст разбивается на предложения, каждое из которых анализируется отдельно по словам. Выделенное слово сопоставляется со списками лингвистической части базы данных. Если первым символом в слове является #, оно сравнивается со списком хэштегов. При совпадении с одним из элементов списка вес хэштега (f_h) добавляется в соответствующий счетчик оценки текста. Если выделенное слово является аспектной категорией либо ключевым словом, его ближайший контекст (одно-два слова) проверяется на наличие в списке оценочных лексических единиц. Если выделенное слово является стоп-словом, дальнейший анализ ближайшего контекста не проводится,

а предложение анализируется на наличие ключевых слов. Если в предложении нет ключевых слов, оно проверяется на наличие оценочных единиц, вес которых будет добавлен в соответствующий счетчик. Каждое направление оценки имеет похожие циклы проверки, отличающиеся только конечным результатом. Во-первых, проходит проверка на наличие в предложении слов-интенсификаторов. Каждая ветвь проверки на наличие интенсификаторов заканчивается проверкой на наличие перед словом слов-инверторов. Если перед оценочным словом есть инвертор, но нет интенсификатора, то в соответствии со шкалой оценки его вес меняется на противоположный и добавляется в соответствующий счетчик. Если перед оценочным словом есть как интенсификатор, так и инвертор, то в счетчик добавляется новое значение веса без изменения направления оценки.

На третьем этапе компьютер суммирует показатели счетчиков для каждой аспектной категории в единый счетчик веса положительной или отрицательной оценки текста твита. Сравнивая счетчики положительных и отрицательных выражений, система делает вывод о направлении оценки всего текста. Если во временной памяти находятся какие-либо значения, они добавляются к общему весу оценки твита.

На четвертом этапе после анализа всех твитов компьютер суммирует показатели всех счетчиков веса положительной и отрицательной оценки, а после их сравнения делает окончательный вывод о репутации личности политического лидера.

Проиллюстрируем адекватность работы формальной модели на массиве текстов десяти твитов. Семь следующих твитов были взяты из материала исследования:

1. *Replying to @realDonaldTrump*

Trump is disrespecting USA. He is a disgrace to the office of the Presidency, for so many reasons. All the lies and distractions DO NOT fool the politically educated. FOX can attempt to cover Trump by distraction, but it will not work!

2. *Replying to @realDonaldTrump*

You are A GREAT PRESIDENT MR. TRUMP!!! Keep up The GOOD WORK!!! Standing With YOU All The Way!!! 🇺🇸👍👍👍🇺🇸

3. *Donald Trump is the worst fake president ever.*

4. *Finally people can focus on Donald Trump's totally decent and competent and sane presidency.*

5. *Replying to @realDonaldTrump*

I'm truly grateful that Trump was elected President instead of Hillary. President saved America.

6. *Replying to @realDonaldTrump*

God bless President Trump the best president ever of the United States #Trump2020

7. *Replying to @realDonaldTrump*

DUMB DUMB DONALD Future generations will know that you and your administration colluded & obstructed justice. The will come out that you are a self serveing low life criminal. You are a useless & inept president. Unable to organise anything.

Из социальной сети Twitter было дополнительно взято еще три твита:

8. *Replying to @realDonaldTrump*

You're losing it. You're a delusional moron. #ImpeachDonaldTrumpNOW

9. *Replying to @realDonaldTrump*

You just keep doing the great job you have been since inauguration! Your supporters don't believe anything your haters say and reasonable people won't either. #Trump2020

10. *Replying to @realDonaldTrump*

He gets more delusional each and every day. Such a sad excuse of a man.

В твите № 1 нормализация проводится не будет. Также не будет обнаружено эмоджи и сочетаний более одного знака препинания. Однако в конце последнего предложения текста стоит восклицательный знак, поэтому во временную память f_{mi} будет занесено значение **0.5**. Анализируя первое предложение, система найдет слово *Trump* в списке ключевых слов, а слово *disrespecting* — в списке отрицательно окрашенной лексики с оценочным весом -1 . Следова-

тельно, счетчик для негативных единиц первой аспектной категории *D. Trump as a president* будет иметь вид $k_{n1} = -1$. Во втором предложении есть два аспектных термина — *office* и *Presidency*, а также одно негативное оценочное слово *disgrace* (-2), которое относится к слову *office*, поэтому значение счетчика для отрицательных оценочных единиц k_{n3} аспектной категории *D. Trump's administration* равно -2. В третьем предложении нет ключевых слов, поэтому вес слов с отрицательной оценочной семантикой *lies* (-1) и *distractions* (-1) вместе с интенсификатором *all* (1.25) добавляется в общий счетчик $K_n = -3.25$. Перед словом *fool* (-1) стоит инвертор *not*, поэтому $K_p = 1$. Далее следует положительно окрашенное слово *educated* (+1), вес которого добавляется в счетчик для положительных единиц твита: $K_p = 1 + 1 = 2$. В последнем предложении за ключевым словом *Trump* следует негативная лексическая единица *distraction* (-1), следовательно, $k_{n1} = (-1) + (-1) = -2$. В последнем предложении есть инвертор *not*, но после него нет никакой эмоционально окрашенной лексики. Значения счетчиков отдельных аспектных категорий суммируются и результаты попадают в соответствующие общие счетчики всего твита. В результате $K_p = 2$ и $K_n = -7.25$. На основании этой информации компьютер делает вывод о том, что направление оценки твита негативное, поэтому из временной памяти f_{mi} в общий счетчик отрицательных единиц добавляется оценочный вес восклицательного знака $K_n = -7.25 + (-0.5) = -7.75$. Таким образом, в тексте твита выражена свехотрицательная оценка Д. Трампа. Система запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

В твите № 2 нормализация лексических единиц не проводится. В данном тексте система обнаружит три положительных эмоджи с общим весом 2.5 (K_p) и три одинаковых эмоджи-интенсификатора ($f_{ei} = 1.5$). Все три предложения твита заканчиваются сочета-

ниями из трех восклицательных знаков, поэтому $f_{mi} = 6$. В первом предложении есть два ключевых слова *PRESIDENT* и *TRUMP* и одна эмоционально окрашенная лексическая единица *GREAT* (+1), которая стоит перед словом *PRESIDENT*. Следовательно, в положительный счетчик k_{p2} аспектной категории *D. Trump as a president* компьютер занесет 1. Так как в оставшихся предложениях отсутствуют и эмоционально окрашенные, и ключевые слова, то система суммирует значения всех промежуточных счетчиков $2.5 + 1 = 3.5$ и помещает результат в общий счетчик K_p . Направление оценки твита положительное, следовательно, из временной памяти f_{ei} и f_{mi} добавляются соответствующие значения и общий счетчик $K_p = 3.5 + 1.5 + 6 = 11$. На основании окончательного результата система делает вывод, что оценка Д. Трампа в твите свехположительная. Компьютер запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

Твит № 3 состоит из одного заканчивающегося точкой предложения, в котором нет эмоджи, сочетаний знаков препинания и ненормализованной лексики. В ходе анализа система обнаружит, что за аспектным термином *Donald Trump* стоит негативно окрашенное слово *worst* (-2), перед ключевым словом *president* есть лексическая единица *fake* (-1), а за ней — интенсификатор с высокой степенью усиления оценки *ever* (1.25). В результате $k_{n1} = -2$, $k_{n2} = (-1) + (-1.25) = -2.25$. Далее компьютер определит, что общий счетчик для негативных единиц твита K_n равен -4.25, т. е. в тексте твита выражена отрицательная оценка Д. Трампа. Система запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

В твите № 4 система определит, что к аспектному термину *Donald Trump's* относится положительное слово *decent* (+1) с интенсификаторами *finally* и *totally*, а к ключевому слову *presidency* — положительные

слова *competent* (+1) и *sane* (+1), хотя этот вывод ошибочный. На общий вывод об оценочности твита это никак не повлияет, но в счетчики по отдельным аспектным категориям будут внесены неверные значения. Итоговые значения счетчиков: $k_{p1} = 2.75$, $k_{p2} = 2$, $K_p = 4.75$. Следовательно, система делает вывод о положительной оценке Д. Трампа. Компьютер запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

В первом предложении **твита № 5** есть аспектный термин *Trump*, к которому относится положительное слово *grateful* (+1) с интенсификатором *truly* (0.5), а также аспектный термин *President* и стоп-слово *Hillary*. Во втором предложении за аспектным термином *President* следует слово *saved* (+1). Несмотря на наличие стоп-слова, в результате получится $k_{p1} = 1.5$ и $k_{p2} = 1$. Значение общего счетчика для положительных единиц твита K_p равно **2.5**, поэтому оценка Д. Трампа в твите положительная. Система запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

В тексте **твита № 6** отсутствуют знаки препинания и эмоджи. Система обнаружит хэштег *#Trump2020*, который находится в списке положительных хэштегов и имеет оценочный вес +1. В общий счетчик для положительных единиц добавляется вес этого хэштега, т. е. $K_p = 1$. Далее компьютер обнаружит два ключевых слова *President Trump* и *president*, которые принадлежат одной аспектной категории *Donald Trump as the President of the USA*, а также три слова со сверхположительной оценкой: *God* (+2), *bles* (+2), *best* (+2) и интенсификатор *ever* (1.25). Таким образом, $k_{p2} = 7.25$, $K_p = 8.25$, поэтому система делает вывод о сверхположительной оценке Д. Трампа в данном твите. Компьютер запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

Текст **твита № 7** нуждается в нормализации таких лексических единиц, как *self serving* и *organise*. В результате анализа система обнаружит эмоционально окрашенное слово *DUMB* (-1), употребленное дважды перед аспектным термином *DONALD*. За аспектным термином *administration* стоят негативно окрашенные слова *colluded* (-1) и *obstructed* (-1), а также отрицательные лексические единицы *self-serving* (-1), *criminal* (-2), *unable* (-1) и относящиеся к ключевому слову *president* слова *useless* (-1), *inept* (-1). Итоговые значения счетчиков: $k_{n1} = -2$, $k_{n2} = -2$, $k_{n3} = -2$, $K_n = (-4) + (-2) + (-2) + (-2) = -10$, поэтому оценка Д. Трампа в твите сверхотрицательная. Система запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

В **твите № 8** обнаружен хэштег *#ImpeachDonaldTrumpNOW*, которого нет в списке хэштегов с отрицательной оценкой. Следовательно, система не прибавит вес хэштега в общий счетчик для отрицательных единиц. Также система найдет только одно отрицательно окрашенное слово — *delusional* (-1) и пропустит два других — *losing* и *moron*, что свидетельствует о неполноте базы данных. В результате значение общего счетчика для отрицательных единиц K_n равно **-1**, т. е. в тексте твита выражена отрицательная оценка Д. Трампа. Компьютер запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

В **твите № 9** система сначала занесет во временную память f_{mi} оценочный вес восклицательного знака, равный **0.5**. Далее она обнаружит хэштег *#Trump2020*, который является положительным и имеет вес +1, поэтому общий счетчик для положительных единиц твита K_p будет равен **1**. В первом предложении компьютер найдет положительное слово *great* (+1), а во втором — ключевое слово *supporters*. Значение общего счетчика K_p равно **2**, направление оценки положительное, поэтому из временной памяти f_{mi} к нему добавляется значе-

ние **0.5**, т. е. общий счетчик K_p равен **2.5**. Оценка Д. Трампа в твите положительная. Система запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

Твит № 10 состоит из двух предложений, в которых отсутствуют ключевые слова. Система обнаружит негативно окрашенные слова *delusional* (-1), *sad* (-1), *excuse* (-1) и интенсификатор *more* (1.25), но не учтет интенсификаторы *every*, *each*, *such*, так как они отсутствуют в базе данных. Поэтому итоговое значение счетчика для отрицательных единиц K_n равно **-4.25**, т. е. оценка Д. Трампа в твите отрицательная. Компьютер запоминает значения счетчиков для всего твита, а также отдельных аспектных категорий и переходит к анализу следующего текста.

На четвертом этапе система суммирует значения счетчиков каждой аспектной категории во всех твитах, а также значения счетчиков с общей оценкой каждого твита. Счетчики $k_{p1} = 4.25$, $k_{p2} = 9.25$; счетчики k_{p3} , k_{p4} , k_{p5} будут иметь нулевые значения; $K_p = 30$. Счетчики $k_{n1} = -5$, $k_{n2} = -6.25$; $k_{p3} = -4$; k_{p4} и $k_{p5} = 0$; $K_n = -36.25$. Далее компьютер сравнивает значения указанных счетчиков и делает вывод, что по аспектной категории *D. Trump in general* репутация Д. Трампа отрицательная ($k_1 = -0.75$), по аспектной категории *D. Trump as the President of the USA* — положительная ($k_2 = 5$), по аспектной категории *D. Trump's administration* — отрицательная ($k_3 = -4$). В целом в массиве текстов из двенадцати твитов репутация личности Д. Трампа отрицательная ($K = -36.25 + 30 = -6.25$).

Проверка адекватности работы формальной модели позволила сделать следующие выводы.

1. Компьютер может определить наличие в тексте оценочных слов, их вес и направление оценки (положительная или отрицательная) для каждой присутствующей в тексте твита аспектной категории субъекта *D. Trump*.

2. Компьютер может сделать вполне адекватный вывод об общей оценке текста твита, а после анализа всего массива твитов сформулировать вывод о репутации Д. Трампа.

3. Несмотря на использование в процессе анализа ключевых слов, относящихся к разным аспектным категориям, система не всегда может адекватно определить, к какому ключевому слову относится оценочное слово, что сказывается на результате оценки разных аспектных категорий субъекта, хотя на общую оценку сильно не влияет.

4. Проблема коррекции орфографических ошибок в твите решена путем добавления блока нормализации, который основан на базе данных ненормализованных лексических единиц, встретившихся в материале исследования, а также грамматически правильных вариантов их написания. В то же время пользователи Twitter часто неправильно расставляют знаки препинания или игнорируют их вообще, что препятствует корректному анализу текста твита.

6. Представленные выше база данных и формальная модель разработаны в рамках инженерного подхода к автоматическому анализу текста. В то же время они могут стать основой для создания классификатора нейронной сети, которая будет самообучаться и точнее распознавать оценочные единицы в высказываниях пользователей социальных медиа с целью последующего формулирования на их основе вывода о репутации личности, группы лиц или организации.

СПИСОК ЛИТЕРАТУРЫ

1. Арутюнова Н. Д. Язык и мир человека. М.: Языки русской культуры, 1999. 896 с.
2. Даль В. И. Толковый словарь живого великорусского языка. Т. 2. М.: ОЛМА-ПРЕСС, 2002. 1088 с.
3. Елишина Е. Имидж и репутация: в чем отличия? [Электронный источник]. URL: [http://www.ug.ru / archive/34080](http://www.ug.ru/archive/34080) (дата обращения: 11.10.2019).
4. Ефремова Т. Ф. Мнение. [Электронный источник]. URL: <https://www.efremova.info/word/mnenie.html#.XGfz8KIzbIU> (дата обращения: 14.10.2019).

5. *Кон И. С.* Социология личности. М.: Политиздат, 1967. 383 с.
6. *Крючкова Н. В.* Что такое престиж? [Электронный источник]. URL: <http://naukarus.com/chto-takoe-prestizh> (дата обращения: 14.10.2019).
7. *Леонтьев А. Н.* Потребности, мотивы и эмоции. М.: Изд-во МГУ, 1971. 40 с.
8. *Мальцев О.* Субъективное мнение — что это такое, чем отличается от объективного? Примеры применения. [Электронный источник]. URL: <https://znay.co/178-subektivnoe-mnenie-chto-eto.html> (дата обращения: 3.11.2019).
9. *Матуров З.* Субъективное и объективное мышление. [Электронный источник]. URL: <https://www.proza.ru/2015/10/03/616> (дата обращения: 4.11.2019).
10. *Наер В. Л.* Из лекций по теоретическим основам интерпретации текста. М.: МГЛУ, 2001. 75 с.
11. *Ожегов С. И., Шведова Н. Ю.* Толковый словарь русского языка: 80 000 слов и фразеологических выражений. М.: ООО «А ТЕМП», 2006. 944 с.
12. *Трубецкой А. Ю.* Категория репутации в социально-политической коммуникации: автореф. дис. ... д-ра психол. наук. М., 2006. 48 с.
13. *Устинова Н. В.* Политическая репутация: сущность, особенности, технологии формирования: автореф. дис. ... канд. полит. наук. Екатеринбург, 2005. 18 с.
14. *Ушаков Д. Н.* Толковый словарь современного русского языка. М.: Аделант, 2014. 800 с.
15. *Dingledine R.* Reputation. [Электронный источник]. URL: <http://www.freehaven.net/~arma/jean.html> (дата обращения: 10.11.2019).
16. *Hatzivassiloglou V.* Effects of adjective orientation and gradability on sentence subjectivity // Proceedings of the 18th Conference on computational linguistics. Stroudsburg: Association for Computational Linguistics, 2000. P. 299–305.

REFERENCES

1. *Arutyunova N. D.* Yazyk i mir cheloveka. M.: Yazyki russkoy kul'tury, 1999. 896 s.
2. *Dal' V. I.* Tolkovyj slovar' zhivogo velikoruskogo yazyka. T. 2. M.: OLMA-PRESS, 2002. 1088 s.
3. *Elshina E.* Imidzh i reputatsiya: v chem otlichiya? [Elektronnyj istochnik]. URL: <http://www.ug.ru/archive/34080> (data obrashcheniya: 11.10.2019).
4. *Efremova T. F.* Mnenie. [Elektronnyj istochnik]. URL: <https://www.efremova.info/word/mnenie.html#XGfz8KIzbIU> (data obrashcheniya: 14.10.2019).
5. *Kon I. S.* Sotsiologiya lichnosti. M.: Politizdat, 1967. 383 s.
6. *Kryuchkova N. V.* Chto takoe prestizh? [Elektronnyj istochnik]. URL: <http://naukarus.com/chto-takoe-prestizh> (data obrashcheniya: 14.10.2019).
7. *Leont'ev A. N.* Potrebnosti, motivy i emotsii. M.: Izd-vo MGU, 1971. 40 s.
8. *Mal'tsev O.* Sub"ektivnoe mnenie — chto eto takoe, chem otlichaetsya ot ob"ektivnogo? Primery primeneniya. [Elektronnyj istochnik]. URL: <https://znay.co/178-subektivnoe-mnenie-chto-eto.html> (data obrashcheniya: 3.11.2019).
9. *Maturov Z.* Sub"ektivnoe i ob"ektivnoe myshlenie. [Elektronnyj istochnik]. URL: <https://www.proza.ru/2015/10/03/616> (data obrashcheniya: 4.11.2019).
10. *Naer V. L.* Iz lektsiy po teoreticheskim osnovam interpretatsii teksta. M.: MGLU, 2001. 75 s.
11. *Ozhegov S. I., Shvedova N. Yu.* Tolkovyj slovar' russkogo yazyka: 80 000 slov i frazeologicheskikh vyrazheniy. M.: ООО «А ТЕМП», 2006. 944 s.
12. *Trubetskoy A. Yu.* Kategoriya reputatsii v sotsial'no-politicheskoy kommunikatsii: avtoref. dis. ... d-ra psihol. nauk. M., 2006. 48 s.
13. *Ustinova N. V.* Politicheskaya reputatsiya: sushchnost', osobennosti, tehnologii formirovaniya: avtoref. dis. ... kand. polit. nauk. Ekaterinburg, 2005. 18 s.
14. *Ushakov D. N.* Tolkovyj slovar' sovremennogo russkogo yazyka. M.: Adelant, 2014. 800 s.
15. *Dingledine R.* Reputation. [Elektronnyj istochnik]. URL: <http://www.freehaven.net/~arma/jean.html> (data obrashcheniya: 10.11.2019).
16. *Hatzivassiloglou V.* Effects of adjective orientation and gradability on sentence subjectivity // Proceedings of the 18th Conference on computational linguistics. Stroudsburg: Association for Computational Linguistics, 2000. P. 299–305.