

РАЗРАБОТКА АЛГОРИТМА ОПРЕДЕЛЕНИЯ МНЕНИЙ В МЕДИАДИСКУРСЕ НА АНГЛИЙСКОМ И БЕЛОРУССКОМ ЯЗЫКАХ

В. В. КОЗЛОВА¹⁾

¹⁾Минский государственный лингвистический университет,
ул. Захарова, 21, 220034, г. Минск, Беларусь

Рассматривается принципиальный алгоритм определения и классификации мнений в информационных и аналитических жанрах медийного дискурса на английском и белорусском языках, разработанный на основе словаря языковых единиц, маркирующих высказывания-мнения (733 единицы, функционирующих в более 5000 высказываний-мнений). Преимущество предложенной модели заключается в том, что она учитывает семантику выявленных единиц, их частотность, а также типологию высказываний-мнений. В действия алгоритма включен ряд шагов: поиск маркеров мнения на основе разработанного словаря маркеров; присвоение каждому предложению определенного «веса» в зависимости от обнаруженных маркеров по сумме числовых значений из словаря (данные по «весу» языковых маркеров указываются в обобщенном виде); разделение высказываний на мнения-интерпретации, мнения-теории, мнения-предположения и комбинированные мнения. Классификация также учитывает семантические и количественные показатели.

Ключевые слова: алгоритм; поиск мнений; словарь языковых маркеров; медийный дискурс; автоматическая классификация мнений.

Благодарность. Автор выражает признательность за помощь в выполнении исследования доктору филологических наук, профессору Г. П. Карпилович.

РАСПРАЦОЎКА АЛГАРЫТМУ ВЫЯЎЛЕННЯ МЕРКАВАННЯЎ У МЕДЫЯДЫСКУРСЕ НА АНГЛІЙСКОЙ І БЕЛАРУСКАЙ МОВАХ

В. В. КАЗЛОВА^{1*}

^{1*}Мінскі дзяржаўны лінгвістычны ўніверсітэт,
вул. Захарава, 21, 220034, г. Мінск, Беларусь

Разглядаецца прынцыповы алгарытм ідэнтыфікацыі і класіфікацыі меркаванняў у інфармацыйных і аналітычных жанрах медыйнага дыскурсу на англійскай і беларускай мовах, распрацаваны на аснове слоўніка моўных адзінак, якія маркіруюць выказванні-меркаванні (733 адзінкі, якія функцыянуюць у больш чым 5000 выказваннях-меркаваннях). Перавага прапанаванай мадэлі заключаецца ў тым, што яна ўлічвае семантыку моўных адзінак, іх колькасныя даныя, а таксама тыпалогію выказванняў-меркаванняў. У дзеянні алгарытму ўключаны шэраг крокаў: пошук маркераў меркавання на аснове распрацаванага слоўніка маркераў; прысваенне кожнаму сказу пэўнай «вагі»

Образец цитирования:

Козлова ВВ. Разработка алгоритма определения мнений в медиадискурсе на английском и белорусском языках. *Журнал Белорусского государственного университета. Филология.* 2020;3:40–49.

For citation:

Kazlova VV. The design of the algorithm for opinion mining in the media discourse in English and Belarusian. *Journal of the Belarusian State University. Philology.* 2020;3:40–49. Russian.

Автор:

Вероника Викторовна Козлова – аспирантка кафедры теории и практики английского языка переводческого факультета. Научный руководитель – доктор филологических наук, профессор Е. Г. Задворная.

Author:

Veranika V. Kazlova, postgraduate student at the department of English theory and practice, faculty of translation and interpretation.
ruis1@yandex.ru

ў залежнасці ад выяўленых маркераў па суме лікавых значэнняў са слоўніка (даныя па «вазе» моўных маркераў прыводзяцца ў агульным выглядзе); падзел на меркаванні-інтэрпрэтацыі, меркаванні-тэорыі, меркаванні-прагнозы і камбінаваныя меркаванні. Класіфікацыя таксама ўлічвае семантычныя і колькасныя паказчыкі.

Ключавыя словы: алгарытм; пошук меркаванняў; слоўнік моўных маркераў; медыйны дыскурс; аўтаматычнае класіфікацыя меркаванняў.

Падзяка. Аўтар выказвае падзяку за дапамогу ў выкананні даследавання доктару філалагічных навук, прафесару [Т. П. Карпіловіч].

THE DESIGN OF THE ALGORITHM FOR OPINION MINING IN THE MEDIA DISCOURSE IN ENGLISH AND BELARUSIAN

V. V. KAZLOVA^a

^aMinsk State Linguistic University, 21 Zacharava Street, Minsk 220034, Belarus

The article runs about the design of the algorithm for opinion mining in the media discourse in English and Belarusian. The algorithm is based on the vocabularies of linguistic units that mark opinions in the media discourse (733 units that are mentioned in more than 5000 opinionated contexts). The advantage of the proposed model lies in the fact that it considers the semantical meaning of the linguistic markers in both languages, their frequency, and the typology of opinionated passages in the context of automatic sorting of opinions. The algorithm includes a number of steps: the search of opinion markers based on the developed vocabularies of linguistic markers, the assignment of a certain «weight» to each sentence depending on the markers (by the sum of numeric values from the vocabulary; the article provides data on the «weight» of linguistic markers in a generalized form), then statements are divided into opinions-interpretations, opinions-theories, opinions-assumptions and an opinion's combined type. The classification takes into account semantic and quantitative indicators.

Keywords: algorithm; opinion mining; vocabulary of linguistic markers; media discourse; automatic classification of opinions.

Acknowledgements. The author wishes to present her gratitude to [Т. П. Карпілович], doctor of science (philology), full professor.

Введение

В современной лингвистике множество исследований посвящено различным аспектам языка СМИ. Изучаются проблемы массовой коммуникации, специфика отдельных жанров медийного дискурса и их языковая репрезентация, проблемы функционирования языка в сфере массовой коммуникации (см., например, [1–3]). Начало исследованиям в области медийного дискурса в отечественном языкознании положили работы М. Е. Тикоцкого, А. И. Наркевича, М. П. Карпович, посвященные стилистике белорусского языка в целом и публицистического текста в частности, анализу публицистических жанров, организации художественного текста и выявлению медиалингвистических приоритетов. Современные белорусские исследования медийного дискурса касаются лингвистической организации текста, интертекстуальности, структурных типов эссе, отдельных аспектов, например, метонимии, средств разговорного синтаксиса и т. д.¹ (см., например, [4]). Как отмечает В. И. Ивченков, сегодня мы видим возникновение журналистики мнений, медиацентричность мира [5, с. 85], что делает изучение высказываний-мнений в медийном дискурсе особенно актуальным.

В последние годы как в зарубежном, так и в отечественном языкознании актуальность и активное развитие получает новое направление в сфере автоматической обработки текста – автоматическое выявление мнений и определение тональности какого-либо текста (*opinion mining and sentiment analysis*).

¹Зелянко С. В. Інтэртэкст у публіцыстычным маўленні : аўтарэф. дыс. ... канд. філал. навук : 10.01.10. Мінск, 2012. 18 с. ; Глазко П. П. Структурно-семантичныя тыпы эсэ в медыйном дыскурсе і іх языковыя характарыстыкі (на матэрыяле англійскай і беларускай прэсы) : дыс. ... канд. філал. навук : 10.02.20. Мінск, 2015. 156 с. ; Горбач В. А. Метафізіка ў публіцыстычным маўленні: парадыгматычныя здольнасці і сінтагматычная неабходнасць : дыс. ... канд. філал. навук : 10.01.10. Мінск, 2000. 145 с. ; Гребень Т. Н. Средства разговорного синтаксиса в медыйном дыскурсе (на матэрыяле беларускаго і англійскаго языков) : дыс. ... канд. філал. навук : 10.02.20. Мінск, 2016. 145 с.

Целью данного направления является установление отношения автора текста и отдельных коммуникантов к некоторой теме, проблемному вопросу, продукту, лицу, явлению и т. п. (см., например, [6; 7]).

В рамках компьютерной лингвистики уже был предпринят ряд попыток по созданию алгоритмов и программ на их основе для определения мнений или тональности [8–10]. Тем не менее, как правило, в фокусе подобных исследований остаются пользовательские отзывы и рецензии и, как следствие, не столько мнения, сколько оценки. В связи с этим целесообразным представляется создание отдельных баз знаний (словарей маркеров) для выявления мнений. При этом следует учитывать, что интерпретация и систематизация мнений возможны только в соответствии с определенным жанром или предметной областью. Это позволит составить эффективный алгоритм идентификации мнений для выбранной предметной области, например, для определенных жанров медиадискурса.

В качестве маркеров мнения могут выступать лексические (специфические слова и сочетания), грамматические, стилистические единицы, вербализирующие мыслительные операции: интерпретацию, логические умозаключения, обобщение и т. п. Традиционным объектом исследования по рассматриваемой проблематике служат оценочная и эмоционально окрашенная лексика и глаголы мнения. Вместе с тем комплексный подход к выявлению языковых средств репрезентации мнения (например, модальные слова и выражения, императивные конструкции, вопросно-ответные комплексы, причинно-следственные структуры, цитация, отдельные специфические для конкретных медиажанров устойчивые единицы и т. д.) позволит добиться не достигавшихся ранее результатов по автоматическому поиску, систематизации, интерпретации и обобщению мнений, представленных в медиадискурсе.

Цель настоящей статьи состоит в освещении разработанного принципиального алгоритма определения мнений в медиатекстах информационных и аналитических жанров на белорусском и английском языках на основе составленного в процессе исследования словаря языковых средств выражения мнения.

Теоретические основы

Если обратиться к направлению автоматической обработки мнений в определенном тексте (*opinion mining*, о котором речь шла выше), то в широком смысле текстовую информацию можно разделить на два типа – факты и мнения [11, р. 460]. Под мнениями следует понимать, во-первых, собственно их контексты, а во-вторых, сравнения [11, р. 460–466]. За собственно мнения принимаются фрагменты текстов, в которых зафиксированы языковые единицы, обозначающие эмоции и чувства, а за сравнения – фрагменты текстов, где рассматриваются сходства и различия каких-либо объектов.

В других исследованиях также проводится противопоставление фактов и мнений. При этом отмечается, что знания принимают форму фактов, а субъективная информация, некоторые неverified высказывания фигурируют в области мнений [12]. Следует отметить, что определенные мнения, пройдя процедуру верификации, могут способствовать получению новых знаний и с течением времени перейти в сферу фактов [13, р. 45–46].

Что касается мнений и фактов, которые представлены в СМИ, необходимо понимать процесс написания публицистического текста. В широком смысле любой текст СМИ – это единый контекст мнения, поскольку его пишет журналист, имея при этом ту или иную коммуникативную цель. Соответственно, вся информация проходит отбор. Одни факты в тексте освящаются, другие получают авторскую интерпретацию, третьи не проходят процесс отбора и в окончательном тексте никаким образом не упоминаются. Следовательно, публицистический текст можно рассматривать в качестве одной единицы мнения, в которой автор представляет некоторое субъективное суждение читательской аудитории на основе отбора фактов и мнений согласно выбранной коммуникативной цели высказывания [14, р. 239]. В данном случае контекст мнения выражает ценности и убеждения автора, редакции или заказчика материала (имеются в виду имиджевые статьи).

Рассматривая мнения, которые представлены в СМИ, необходимо отметить проблему разграничения мнений и оценок. Общим является то, что и мнения, и оценки отражают убеждения, воззрения, «чувствования» общности людей, психологический уклад социум [15, р. 28] и представляют информацию об объекте на основе внутренней картины мира того или иного человека [16, с. 7]. В некоторых исследованиях оценка включена в структуру мнения, выступает или его частью, или подвидом. Например, оценка считается высказыванием-мнением относительно ценности какого-либо объекта [17, с. 26; 20, с. 8], а также субъективной истиной, соответствующей картине мира говорящего [18, с. 51]. И для мнений, и для оценок существуют отдельные массивы языковых средств, их выражающих. В других работах отмечается сходство мнений и оценок и в то же время говорится об отсутствии полного совпадения данных областей (речь идет о так называемых экзистенциальных мнениях и оценках-ощущениях) [16, с. 8].

Под мнением в данной работе понимается некоторый фрагмент текста, в котором представлено высказывание об определенном объекте и отмечена одна или несколько языковых единиц, маркирующих мнение.

Для определения тональности текста и установления высказываний мнения наиболее эффективными признаны метод векторного анализа, метод, основывающийся на заранее составленных словарях (определение тональности, языковых маркеров мнения), с применением лингвистического анализа, а также гибридный метод. Векторный анализ предполагает сравнение текстов с размеченным заранее эталонным корпусом текстов, однако среди недостатков метода указывают невозможность установления эмотивности на уровне предложения, а также ограниченную разнородность текстов корпуса. Метод, основывающийся на словарях маркеров, подразумевает определение тональности текста и установление высказываний мнения по совокупности тональной лексики и языковых маркеров мнения, но при этом сложно дать количественную оценку позитивности или негативности анализируемого текста. Гибридный метод включает оба подхода.

В данном исследовании рассматривается разработанный автором алгоритм поиска и анализа мнений, которые фигурируют в медийном дискурсе, основанный на словарях языковых маркеров, поскольку он наиболее эффективен для поиска мнений в медийном дискурсе. Кроме того, именно этот подход позволит применять полученные данные для исследования других текстов информационных и аналитических жанров медийного дискурса. В представленной модели также учитываются количественные показатели.

Результаты и их обсуждение

Материалом анализа послужили 200 текстов различных информационных и аналитических жанров на английском и белорусском языках (мнения и комментарии, аналитические статьи, новостные заметки, репортажи и др.). Источниками практического материала стали как национальные, так и региональные британские и белорусские газеты («The Guardian», «The Daily Telegraph», «The Independent», «The Scotsman», «Evening Standard», «Звязда», «Культура», «Новы час», «Гомельская праўда» и др.).

Методы анализа входят в совокупность исследовательских процедур, целесообразность которых диктуется темой, задачами и материалом исследования и которые включают сравнительно-сопоставительный анализ, контекстуальный метод, контент-анализ, методы моделирования, количественной обработки данных, интерпретации и классификации исследуемого материала.

Прежде всего для составления эффективного алгоритма поиска и распознавания мнений необходимо создать базу данных (словарь) определенных языковых единиц, под которыми мы понимаем маркеры мнения, т. е. языковые средства, благодаря которым тот или иной фрагмент можно маркировать как высказывание-мнение.

В корпусе медийных текстов, отобранных для исследования, были обнаружены разноуровневые языковые маркеры в количестве 733 единиц, функционирующих в более 5000 высказываний-мнений, в структуре которых выделяются:

- 1) единицы морфологического уровня:
 - глаголы (с одной стороны, глаголы, являющиеся маркерами ввода мнения, с другой – глаголы, входящие в структуру собственно мнения);
 - существительные;
 - служебные части речи (было выявлено на английском языке 24 единицы, в белорусском – 17);
 - формы будущего времени и сослагательного наклонения;
- 2) единицы лексического уровня:
 - лексемы единицы с модальным значением (в английском языке – 20, в белорусском – 10);
- 3) единицы синтаксического уровня:
 - вопросительные предложения (запрос информации; вопрос, задающий проблемное поле; риторический вопрос; вопрос, в котором предлагаются идеи; вопросно-ответный комплекс; коммуникативный вопрос; вопрос-сомнение; альтернативный вопрос)
 - комментарии (в английском языке – 62, в белорусском – 91);
 - императив.

В результате систематизации материала был составлен словарь маркеров (табл. 1).

Качественная журналистика предполагает исключение авторской позиции при освещении информационных поводов, поэтому профессиональные журналисты стремятся раскрыть содержание новости объективно. Однако мнения являются неотъемлемой частью газетных материалов. В значительном количестве мнения функционируют в текстах передовых статей, авторских колонок, в аналитических материалах, отзывах и рецензиях, даже в новостных статьях и во многих других типах медиатекстов. Нередко возникают трудности при разграничении фактов и мнений, надежной верифицируемой информации и частной точкой зрения.

Таблица 1

Фрагмент сводной таблицы маркеров мнения на английском и белорусском языках

Table 1

Fragment of the summary table of opinion markers in English and Belarusian

Группы маркеров мнения	Данные для английского языка		Данные для белорусского языка	
	Количество единиц	Примеры	Количество единиц	Примеры
<i>Глаголы, входящие в структуру мнения</i>				
Выражение причинно-следственных отношений, логических взаимосвязей	12	to affect, to cause, to lead to, etc.	16	аказацца, сведчыць пра і інш.
Выражение эмоций и чувств	3	to feel, to express condolences, etc.	8	выклікаць (здзіўленне, абурэнне, бурныя эмоцыі і г. д.) і інш.
Выражение дополнительных чувств (маркеры согласия/несогласия)	28	to accuse, to blame, to criticize, etc.	10	быць супраць, мець рацыю, пагаджацца і інш.
Выражение желаний, идей, целей	12	to aim at, to desire, to hope, etc.	9	марыць, прапанаваць, хацець і інш.
Представление дефиниционных отношений	14	to represent, to mean, to refer to	4	з'яўляцца, лічыцца і інш.
Маркеры с нейтральным значением	10	to acknowledge, to confirm, etc.	9	меркаваць, ставіцца і інш.
Выражение прогностических значений	2	to plan, to expect, etc.	1	планаваць і інш.
<i>Глаголы ввода мнения в предложение</i>				
Глаголы говорения	18	to say, to ask, to quote, etc.	13	сказаць, запытацца, дадаць і інш.
Описание поведения говорящего	1	to sing, etc.	5	уздыхнуць, усміхнуцца і інш.
Представление логических взаимосвязей	11	to compare, to conclude, to estimate, etc.	8	падсумаваць, разважаць і інш.
Маркеры с нейтральным значением	29	to think, to publish, to mention, etc.	10	думаць, меркаваць і інш.
Представление эмоционального состояния говорящего	5	to blame, to feel, etc.	15	дзяліцца крыўдай, абнадзейваць, буркатаць і інш.
Выражение степени уверенности	5	to point out, to suppose, etc.	10	падкрэсліць, звярнуць увагу і інш.
Выражение желаний, идей, целей	2	to aim at, to advise, etc.	5	паабяцаць, прапанаваць і інш.
<i>Существительные</i>				
Указание на источник информации	20	rating, Twitter, etc.	9	гутарка, даследаванне і інш.
Маркеры с нейтральным значением	20	question, comment, etc.	17	гутарка, тэзіс, меркаванне і інш.
Выражение желаний, идей, целей	17	desire, problem, etc.	16	парада, задача, выхад і інш.
Выражение прогностических значений	2	trend, tendency, etc.	6	прагноз, тэндэнцыя і інш.
Представление дефиниционных отношений	3	sense, truth, etc.	5	праўда, вызначэнне і інш.
Представление реакции коммуникантов	10	outcry, doubt, etc.	5	заклік, уражанне і інш.
Представление логических взаимосвязей	9	argument, impact, etc.	9	выснова, прычыны і інш.

В современных медиатекстах весь объем информации можно разделить на объективную и субъективную. При этом существующие на данный момент поисковые программы сосредоточены на идентификации фактов и не в состоянии выполнить поиск мнений. Фактическую информацию относительно легко автоматически найти по набору тематических ключевых слов, в то время как определить мнения подобным методом не представляется возможным, например, в поисковых запросах типа «Что думают люди об уровне зарплаты?»; «Какие есть предложения по улучшению системы образования в Беларуси?». Это означает, что существующие поисковые системы не справляются с запросами по извлечению мнений из текстов.

Допустим, пользователь намеревается ввести следующий запрос: «Как улучшить школьное образование в Беларуси?». Его интересуют публикации в белорусских СМИ за 2015–2016 гг., он хочет отобрать лишь мнения экспертов и непосредственных участников ситуации (в данном случае – представителей образовательной сферы), мнения журналистов при этом его не волнуют. Пользователю необходимо узнать существующие идеи по решению проблемы, исключая интерпретации и прогнозы.

Для решения поставленной задачи мы предлагаем алгоритм поиска и классификации мнений в медиатекстах (см. рисунок).



Принципиальный алгоритм определения мнений в медиадискурсе на английском и белорусском языках

Basic algorithm for opinion mining in media discourse in English and Belarusian

Прежде всего необходимо отобрать подходящие по теме тексты из СМИ. Для этого можно воспользоваться уже существующими разработками (например, [19; 20]).

Далее для удобства отдельное предложение принимается в качестве высказывания-мнения. Соответственно, отобранные тексты разбиваются на предложения. Сделать это несложно, поскольку предложение имеет четкие формальные признаки: оно отделено знаками пунктуации (точка, вопросительный знак, восклицательный знак и т. п.).

Следующим шагом является разделение предложения на слова, чтобы в дальнейшем было возможно искать маркеры.

Затем осуществляется непосредственный поиск маркеров мнения в каждом предложении. Для этого слова в предложениях сверяются с единицами из составленного словаря маркеров (фрагмент представлен в табл. 1).

После выявления маркеров в высказываниях переходим к этапу классификации. Каждому обнаруженному маркеру, а затем и каждому предложению присваивается определенный «вес» по сумме значений «веса» маркеров (табл. 2), который определяется по сумме числовых значений единиц в словаре (табл. 3). В некоторых случаях отдельные маркеры подгрупп получают большее числовое значение, поскольку, согласно проведенному исследованию, они имеют большую частоту употребления.

Таблица 2

Интерпретация значений «веса» маркеров мнения

Table 2

Interpretation of «weight» values for opinion markers

Значение «веса»	Интерпретация
≤ 2	Низкое
от 3 до 5	Среднее
≥ 6	Высокое

Так, предложения с одним маркером типа комментария получают минимальный «вес», равный 1. Вместе с тем предложению с одним маркером с модальным значением будет присвоен вес 6, т. е. достаточно большой, поскольку, исходя из проведенного исследования, единицы с модальным значением высокочастотны в высказываниях-мнениях, в которых предлагаются идеи (условно их можно называть мнениями-теориями).

Далее отобранные предложения разделяются на мнения-цитаты и авторские мнения. Последние принадлежат авторам собственно медиатекстов, в цитатах же представляются мнения экспертов, лидеров мнений, участников обсуждаемой ситуации. Для этого следует воспользоваться существующими алгоритмами определения цитат в тексте (например, [21; 22]).

Таблица 3

Средние значения «веса» маркеров мнения по типам

Table 3

Average «weights» of opinionative markers in terms of their types

Значение «веса»	Пример
1	Комментарии, глаголы ввода эмоциональных характеристик
2	Большая часть глаголов ввода (глаголы говорения, описание поведения говорящего и т. д.), ряд существительных (обозначение источника информации, реакции коммуникантов и т. д.), ряд вопросов (запрос информации, коммуникативный вопрос) и др.
3	Некоторые глаголы ввода (представление логических взаимосвязей, идей), служебные части речи, собственно выражающие мнение существительные (представление идей, прогнозов и т. д.)
4	Собственно выражающие мнение глаголы (представление логических идей, взаимосвязей, дефиниционные отношения и т. д.), вопросы-идеи, альтернативные вопросы и др.
5	Часть модальных единиц, формы будущего времени (в английском языке) и условного наклонения, императив, ряд вопросов (проблемный вопрос, риторический вопрос и др.)
6	Часть модальных единиц, формы будущего времени (в белорусском языке)

Затем в зависимости от типа представленных маркеров предложения разделяются на мнения-интерпретации, мнения-теории, мнения-предположения и комбинированный тип мнений. Классификация производится на основании маркеров словаря.

Последним шагом является выдача результата по запрашиваемым пользователем критериям. При этом высказывания-мнения отображаются в порядке снижения присвоенного «веса» (табл. 4).

Таблица 4

Интерпретация значений «веса» высказывания-мнения

Table 4

Interpretation of «weight» values for opinionative phrases

Значение «веса»	Интерпретация
≤ 3	Низкое
от 4 до 6	Среднее
≥ 7	Высокое

Таким образом, принципиальный алгоритм не только поможет выявить высказывания-мнения в тексте, но классифицировать их определенным образом по требованиям пользователя.

Заключение

Проведенное исследование показало, что в медиадискурсе на белорусском и английском языках представлено значительное количество языковых средств выражения мнения, с помощью которых возможно идентифицировать высказывания-мнения в конкретном тексте, поскольку в большинстве случаев они содержат выраженное эксплицитно отношение коммуниканта к обсуждаемой проблеме, ситуации, явлению и т. д. С учетом высокого содержания маркеров мнения в отобранных медиатекстах целесообразно использовать метод, основанный на заранее составленных тональных словарях языковых единиц, передающих мнение, с применением лингвистического анализа. Языковые средства выражения мнения представлены широким диапазоном маркеров мнения на морфологическом, лексическом, синтаксическом уровнях языковой системы. В результате был составлен принципиальный алгоритм поиска мнений в медиатекстах на белорусском и английском языках, который основывается на разработанном словаре маркеров мнения, учитывается семантика выявленных единиц, их частотность, тип высказываний-мнений.

До настоящего времени исследования маркеров мнения в сопоставительном аспекте (белорусский – английский языки) с целью разработать алгоритм поиска мнений в медийных текстах не предпринимались. Несмотря на некоторые имеющиеся ограничения представленной алгоритмической системы, она позволит получать более точные результаты в процессе поиска и классификации мнений в медиатекстах.

Результаты исследования могут быть инкорпорированы в общий алгоритм поиска мнений и определения тональности в медийных текстах на белорусском и английском языках. Разработанный алгоритм применим прежде всего для анализа СМИ и, соответственно, компьютерных моделей исследования мнения групп людей или отдельных лиц по определенным вопросам (автоматическая оценка деятельности организаций, контроль и оценка эффективности маркетинговых и рекламных проектов, компьютерные системы выбора целевой аудитории, экономические исследования по определению отношения целевой аудитории к какому-либо продукту или сервису, автоматизированные прогностические исследования экономических проблем, автоматическое определение мнения экспертов, оптимизация опросов и т. д.). Поскольку сегодня высказывания по большому количеству популярных тем уже имеются в интернете, предлагаемый алгоритм позволит значительно упростить анализ эффективности той или иной деятельности, избавив от необходимости проведения классического опроса.

Библиографические ссылки

1. Добросклонская ТГ. *Медиалингвистика: системный подход к изучению языка СМИ: современная английская медиаречь*. Москва: Флинта; 2008. 263 с.
2. Красноярова ОВ. Медийный текст: его особенности и виды. *Известия Иркутской государственной экономической академии*. 2010;3:177–181.
3. Тырыгина ВА. *Жанровая стратификация масс-медийного дискурса*. Москва: Либроком; 2010. 320 с.
4. Ивченко ВИ. *Лингвистическая организация текста: в творческой лаборатории Владимира Короткевича*. Минск: Белорусский государственный университет; 2002. 211 с.

5. Ивченко В.И. Журналистика в системе культуры: медиацентричность современного мира. У: Лакотка А.И., редактор. *Першы міжнародны навуковы кангрэс беларускай культуры. Зборнік матэрыялаў; 5–6 мая 2016 г.; Мінск, Беларусь*. Мінск: Права і эканоміка; 2016. с. 278–282.
6. Зализняк А.А. *Многозначность в языке и способы ее представления*. Москва: Языки славянских культур; 2006. 672 с.
7. Анненкова И.В. Язык современных СМИ как система интерпретации в контексте русской культуры. В: Солганик Г.Я., составитель. *Язык современной публицистики*. Москва: Флинта; 2007. с. 99–114.
8. Bing Liu. *Sentiment analysis and opinion mining*. California: Morgan & Claypool Publishers; 2012. 168 p.
9. Pang Bo. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. 2008;2(1–2):1–135. DOI: 10.1561/15000000011.
10. Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings 40th annual meeting of the association for computational linguistics*. Philadelphia: Association for Computer Linguistics; 2002. p. 417–424. DOI: 10.3115/1073083.1073153.
11. Periakaruppan S, Shanmugasundaram H, Joan L. Research directions, challenges and issues in opinion mining. *International Journal of Advanced Science and Technology*. 2013;60:1–8. DOI:10.14257/ijast.2013.60.01.
12. Котельников Е.В., Клековкина М.В. Автоматический анализ тональности текстов на основе методов машинного обучения. *Компьютерная лингвистика и интеллектуальные технологии*. 2012;11:27–36.
13. Bing Liu. *Sentiment analysis and opinion mining*. Luxembourg: Springer Science + Business Media; 2011. 622 p.
14. Tsytsarau M, Palpanas T. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*. 2012;24(3):478–514. DOI:10.1007/s10618-011-0238-6.
15. Edwards D. *Newspeak in the 21st century*. London: Pluto Press; 2009. 304 p.
16. Симонов А.К., Горбаневский М.В. Понятие чести, достоинства и деловой репутации: спорные тексты СМИ и проблемы их анализа и оценки юристами и лингвистами. Москва: Медия; 2004. 328 с.
17. Солошенко Н., Орлова Ю.А., Розалиев В.Л. Автоматическое выделение сюжетов и тем из потока новостных сообщений. *Известия Волгоградского государственного технического университета*. 2015;2:83–90.
18. Гибатова Г.Ф. Об истинности мнений. *Вестник Южно-Уральского государственного университета*. 2011;22:54–58.
19. Иванова Г.Ф. О мнениях и оценках. *Известия Российского государственного педагогического университета имени А. И. Герцена*. 2007;8:25–31.
20. Дмитриевская М.А. Знание и мнение: образ мира, образ человека. В: Арутюнова Н.Д., редактор. *Логический анализ языка. Знание и мнение*. Москва: Наука; 1988. с. 6–17.
21. Горошко Е.И. *Языковое сознание: гендерная парадигма*. Харьков: ИД ИНЖЭК; 2003. 437 с.
22. Hai-Jew S, editor. *Data analytics in digital humanities, multimedia systems and applications*. Manhattan: Springer Science; 2017. 317 p.
23. Poulquien B, Steinberger R, Best C. Automatic detection of quotations in multilingual news. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP'2007)*. Bulgaria: Borovets; 2007. p. 487–492.
24. Clergerie É de L, Sagot B, Stern R, Denis P, Recourcé G, Mignot V. Extracting and visualizing quotations from news wires. In: Vetulani Z, editor. *Human language technology. Challenges for computer science and linguistics, lecture notes in computer science*. Berlin: Springer; 2011. p. 522–532.

References

1. Dobrosklonskaya TG. *Medialingvistika: sistemnyi podkhod k izucheniyu yazyka SMI: sovremennaya angliiskaya mediarech'* [Media linguistics: system approach to the media language studies: modern English media speech]. Moscow: Flinta; 2008. 263 p. Russian.
2. Krasnoyaroova OV. Media text: its characteristics and types. *Izvestiya Irkutskoi gosudarstvennoi ekonomicheskoi akademii*. 2010;3:177–181. Russian.
3. Tyrygina VA. *Zhanrovaya stratifikatsiya mass-mediinogo diskursa* [Genre stratification of the mass media discourse]. Moscow: Librolom; 2010. 320 p. Russian.
4. Ivchenkov VI. *Lingvisticheskaya organizatsiya teksta: v tvorcheskoi laboratorii Vladimira Korotkevicha* [Linguistic organization of the text: in the creative laboratory of Vladimir Korotkevich]. Minsk: Belarusian State University; 2002. 211 p. Russian.
5. Ivchenkov VI. [Journalism in the cultural system: media centrality of the modern world]. In: Lokotko AI, editor. *Pershy mizhnarodny navukovy kangrjes belaruskaj kul'tury. Zbornik matjeryjalaw; 5–6 maja 2016 g; Minsk, Belarus'* [The first international scientific congress of the Belarusian culture: collection of materials; 2016 May 5–6; Minsk, Belarus]. Minsk: Pravo i ekonomika; 2016. p. 278–282. Russian.
6. Zaliznyak AA. *Mnogoznachnost' v yazyke i sposoby ee predstavleniya* [Polysemy in the language and the means of its representation]. Moscow: Yazyki slavyanskikh kul'tur; 2006. 672 p. Russian.
7. Annenkov IV. [The language of the modern mass media as the system of interpretation in the context of Russian culture]. In: Solganic GYa, editor. *Yazyk sovremennoi publitsistiki* [The language of modern journalism]. Moscow: Flinta; 2007. p. 99–114. Russian.
8. Bing Liu. *Sentiment Analysis and Opinion Mining*. California: Morgan & Claypool Publishers; 2012. 168 p.
9. Pang Bo. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. 2008;2(1–2):1–135. DOI: 10.1561/15000000011.
10. Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings 40th annual meeting of the association for computational linguistics*. Philadelphia: Association for Computer Linguistics; 2002. p. 417–424. DOI: 10.3115/1073083.1073153.
11. Periakaruppan S, Shanmugasundaram H, Joan L. Research directions, challenges and issues in opinion mining. *International Journal of Advanced Science and Technology*. 2013;60:1–8. DOI:10.14257/ijast.2013.60.01.
12. Kotelnikov EV, Klekovkina MV. Sentiment analysis of texts based on machine learning methods. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii*. 2012;11:27–36. Russian.

13. Bing Liu. *Web data mining*. Luxembourg: Springer Science + Business Media; 2011. 622 p.
14. Tsytarau M, Palpanas T. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*. 2012;24(3): 478–514. DOI: 10.1007/s10618-011-0238-6.
15. Edwards D. *Newspeak in the 21st century*. London: Pluto Press; 2009. 304 p.
16. Simonov AK, Gorbanevskij MV. *Ponyatie chesti, dostoinstva i delovoi reputatsii: spornye teksty SMI i problemy ikh analiza i otsenki yuristami i lingvistami* [The concept of honor, dignity and business reputation: controversial media texts and problems of their analysis and evaluation by lawyers and linguists]. Moscow: Medeya; 2004. 328 p. Russian.
17. Soloshenko N, Orlova YA, Rosaliev VL. Automatic allocation of subjects and themes from the stream of news reports. *Izvestiya Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta*. 2015;2:83–90. Russian.
18. Gibatova GF. Truth in opinion contexts. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta*. 2011;22:54–58. Russian.
19. Ivanova GF. On opinions and evaluation. *Izvestia: Herzen University Journal of Humanities & Sciences*. 2007;8:25–31. Russian.
20. Dmitrovskaya MA. [Knowledge and opinion: the image of the world, the image of a person]. In: Arutyanova ND, editor. *Logicheskii analiz yazyka. Znanie i mnenie* [Logical analysis of the language, knowledge and opinion]. Moscow: Nauka; 1998. p. 6–17. Russian.
21. Goroshko EI. *Yazykovoe soznanie: gendernaya paradigma* [Language consciousness: a gender paradigm]. Kharkov: ID INJEK; 2003. 437 p. Russian.
22. Hai-Jew S, editor. *Data analytics in digital humanities, multimedia systems and applications*. Manhattan: Springer Science; 2017. 317 p.
23. Pouliquen B, Steinberger R, Best C. Automatic detection of quotations in multilingual news. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP'2007)*. Bulgaria: Borovets; 2007. p. 487–492.
24. Clergerie É. L, Sagot B, Stern R, Denis P, Recourcé G, Mignot V. Extracting and visualizing quotations from news wires. In: Vetulani Z, editor. *Human language technology. Challenges for computer science and linguistics, lecture notes in computer science*. Berlin: Springer; 2011. p. 522–532.

Статья поступила в редколлегию 06.10.2020.
Received by editorial board 06.10.2020.