

УДК 004.89:070.15

DOI 10.52575/2712-746X-2024-51-1-201-210

Применение автоматизированных методов и алгоритмов для идентификации идеологем новостных текстов

Секерин А.В., Кудинов В.А.

Курский государственный университет,
Россия, 305000, Курская область, г. Курск, ул. Радищева, д. 33
E-mail: alexei.sekerin@yandex.ru, kudinovva@yandex.ru

Аннотация. На современном этапе развития информационного противоборства особое значение приобретает проблема автоматизации процесса категоризации новостных сообщений по центрам влияния, в том числе выявление сообщений, содержащих деструктивные ценностные установки. Несмотря на большое количество публикаций, посвященных описанию методов и технологиям современных информационных войн, лингвистическим аспектам функционирования идеологем, в настоящий момент отсутствуют комплексный подход к автоматизации процесса идентификации идеологем русскоязычных новостных текстов. Целью данного исследования является определение возможности применения автоматизированных методов идентификации идеологем и построение теоретической модели интеллектуальной системы. В результате исследования были проанализированы статистические, нейросетевые и граф-ориентированные подходы, оценена их применимость для идентификации идеологем русскоязычного дискурса посредством искусственных нейронных сетей, а также разработана модель интеллектуальной системы.

Ключевые слова: обработка естественных языков, нейронная сеть, целевой анализ настроений, распознавание именованных сущностей, трансформер, тезаурус

Для цитирования: Секерин А.В., Кудинов В.А. 2024. Применение автоматизированных методов и алгоритмов для идентификации идеологем новостных текстов. Экономика. Информатика. 51(1): 201–210. DOI 10.52575/2712-746X-2024-51-1-201-210

Usage of Automated Methods and Algorithms for Identification of Ideologemes of News Texts

Alexei V. Sekerin, Vitaly A. Kudinov

Kursk State University
33 Radishcheva St, Kursk, Kursk region, 305000, Russia
E-mail: alexei.sekerin@yandex.ru, kudinovva@yandex.ru

Abstract. At the present stage of the development of information warfare, the problem of automating the process of categorizing news texts by centers of influence, including the identification of messages containing destructive value orientations, is of particular importance. Despite the large number of publications devoted to the description of methods and technologies of modern information wars, linguistic aspects of the functioning of ideologues, there is currently no comprehensive approach to automating the process of identification of ideologues of Russian-language news texts. The purpose of this study is to determine the possibility of applying automated methods identification of ideologies and to build a theoretical model of an intelligent system. As a result of the research, statistical, neural network and graph-oriented approaches were analyzed, their applicability to identify ideologemes of Russian-language discourse through artificial neural networks was evaluated, and an intelligent system model was developed.

Keywords: natural language processing, neural network, targeted sentiment analysis, named-entity recognition, transformer, thesaurus

For citation: Sekerin A.V., Kudinov V.A. 2024. Usage of automated methods and algorithms for identification of ideologemes of news texts. Economics. Information Technologies. 51(1): 201–210. DOI 10.52575/2712-746X-2024-51-1-201-210

Введение

Современное информационное противоборство характеризуется активным применением массовых коммуникативных практик посредством специфического употребления единиц языка с целью формирования у общности людей (географической, конфессиональной, политической, экономической) заданного мировоззрения (обобщенной системы взглядов на окружающий мир, роль и место в нем человека, а также соответствующих этому идеалов и убеждений), при одновременном обеспечении безопасности актора.

Идеология как теоретически оформленная система идей, выражающая суть социальных интересов различных групп общества, представляет собой рациональный уровень восприятия действительности.

Увеличение числа ежедневно публикуемых новостных публикаций в сети Интернет, содержащих, в том числе, деструктивные ценностные установки различных центров информационного внимания, а также невозможность применения исключительно требующей значительное количество ресурсных и временных затрат экспертной оценки для обеспечения информационно-психологической безопасности населения свидетельствуют о необходимости разработки интеллектуальных систем анализа идеологем как элементов, отражающих политические идеологии и религиозные учения.

Вопрос функционирования идеологем является предметом исследований в сфере лингвистики, политических и компьютерных наук, теории информации и информационных процессов. Учеными применяются несколько подходов к их выявлению в текстах, один из которых базируется на анализе семантической структуры информационного сообщения. Идеологически окрашенные слова и выражения, а также формулировки, вызывающие эмоциональную реакцию, могут указывать на присутствие идеологем.

В диссертационном исследовании С.А. Журавлева идеологема определена как «знаковое образование идеологического метауровня и дискурсивная единица, значимость которой определяется метаконтекстуально» [Журавлев, 2004].

А.П. Чудиновым под «идеологемой» предложено понимать, в первую очередь, слова и словосочетания, значения которых «неодинаково понимаются сторонниками разных политических взглядов», что находит отражение в эмоциональной окраске слова, «на которое переносится оценка соответствующего явления», а во-вторых, слова и словосочетания, используемые «только сторонниками определенных политических взглядов» [Чудинов, 2012].

А.А. Карамовой предложена классификация идеологем, широко применяемая в настоящее время отечественными лингвистами [Карамова, 2015]:

- лексика и фразеологизмы с идеологическим макрокомпонентом (собственно идеологемы), включающие термины, хотя бы одно из лексических значений которых содержат идеологический компонент;
- идеологизированная лексика – слова и словосочетания, денотация которых осложнена идеологическим микрокомпонентом, отражающим субъективное отношение определенной идеологии к обозначаемому понятию в виде квалификации «свой/чужой»;
- смысловые (контекстуальные, контекстные) идеологемы, идеологическое значение которых не отражается в самой структуре, а формируется контекстом употребления.

В рамках настоящего исследования под идеологемой понимается лексико-фразеологическая единица (слово или словосочетание), отражающая политическую или религиозную идеологию и имеющую высокую степень оценочности.

Целью настоящего исследования является определение возможности применения автоматизированных методов идентификации идеологем и построение теоретической модели интеллектуальной системы.

Для достижения цели исследования необходимо выполнить следующие задачи:

1. Провести анализ методов и алгоритмов классификации текстов по идеологическому принципу и оценить возможность их применения для идентификации идеологем.
2. Разработать модель интеллектуальной системы идентификации идеологем.

Новизна проведенного исследования состоит в обосновании актуальности проблемы автоматической идентификации идеологем и её постановки с помощью расширения существующей проблемы нейросетевой классификации текстов по идеологическому принципу.

Теоретическое применение полученных результатов состоит в установлении методов и алгоритмов, применяемых в процессах автоматизированного анализа идеологических категорий текста.

Практическое значение полученных результатов состоит в возможности использования разработанной теоретической модели для построения интеллектуальных систем идентификации идеологем с использованием искусственных нейронных сетей.

Анализ подходов к автоматизированной идентификации идеологем

В настоящее время в отечественной и зарубежной лингвистике развиты несколько подходов к применению компьютерных методов для автоматизации процесса идентификации идеологем, которые могут быть разделены на 2 группы:

- статистические;
- на основе моделей машинного обучения.

Статистические критерии успешно используются для выявления смысловых (контекстуальных) идеологем в корпусах русскоязычных текстов. В исследовании В.А. Яцко [Яцко, 2016] на основе метода построения таблиц сопряженности («таблиц совместного распределения частот двух и более номинативных признаков, измененных на группе объектов» [Наследов, 2004]) значение критерия хи-квадрат интерпретируется как степень зависимости между данным словом и исследуемым корпусом текстов. Автор делает вывод об отнесении данного слова к категории «идеологема», если вес хи-квадрат термина критическое значение Q , определяемое на основе количества степеней свободы $df = 1$ и уровня значимости $\alpha = 0.05$.

В работах Н.К. Радиной предлагается подход, учитывающий не только разницу частотности слова в исследуемом корпусе относительно эталонного, но частоту употребления идеологемы в тексте и его эмоциональную окраску и оценочность [Радина и др., 2018]. Для каждого слова в изучаемой коллекции вычисляется оценка ipm – число его упоминаний в тексте длиной миллион слов. Результаты сравниваются с эталонными значениями частот в Национальном корпусе русского языка. Выполняется ранжирование кандидатов, у которых $ipm > ipm_{эталон}$, с учетом статистической метрики TF-IDF с целью выявления частотных слов конкретного текста. Однако авторы не уточняют, какое значение метрики является критическим для отбора и дальнейшего экспертного анализа.

$$tfidf_{t,D} = \frac{n_{t,D}}{\sum_k n_{k,D}} \times \log \frac{|D|}{|\{d_i \in D \mid t_i \in d_i\}|},$$

где

$n_{t,D}$ – число упоминаний слова t в документе;

$\sum_k n_{k,D}$ – число слов в документе;

$|D|$ – количество документов в исследуемой коллекции;

$\{|d_i \in D | t_i \in d_i\}$ – количество документов из коллекции D , где слово t встречается хотя бы единожды.

Смежной с задачей идентификации идеологием является проблема определения экстремистских текстов, решения которой включают системы автоматического поиска по ключевым словам и фрагментам текста, отражающим деструктивные политические идеологии. Развитие данного направления в отечественной науке связано с законодательными ограничениями на распространение экстремистских материалов.

В работах группы М.И. Петровского предложен двухэтапный алгоритм выявления экстремистских текстов, включающий выявление ключевых фрагментов посредством латентно-семантического анализа (ЛСА) [Машечкин и др., 2016] и проверки найденных по шаблону, составленному на основе набора Ansar1 [Petrovskiy et al., 2017].

Задача идентификации идеологием может рассматриваться как приложение к решению проблемы автоматической классификации текстов по политическим идеологиям посредством применения искусственных нейронных сетей. В зарубежной литературе преобладает тенденция к разработке моделей глубокого обучения, входными данными которых является текст, а выходными – метка наименования политической идеологии или религиозного течения.

Одним из подходов к интерпретации предсказаний, а именно определение токенов – элементов текста, повлиявших на ответ, является адаптивное предсказание Шепли [Lundberg, 2020] – теоретико-игровой метод, основанный на оценке влияния отдельных признаков примера на прогнозируемое решение. В его основе построение вектора Шепли, то есть такого распределения выигрыша, в котором каждый игрок получает математическое ожидание своего вклада в соответствующие коалиции, определяемое формулой:

$$\Phi(\mathcal{G})_i = \sum_{i=1}^k \frac{(k-1)!(n-k)!}{n!} (\mathcal{G}(K) - \mathcal{G}(K \setminus i)),$$

где:

n – число игроков,

k – количество участников коалиции K .

Применительно к интерпретации языковых моделей под игроками подразумеваются токены (фрагменты) текста, а коалициями – прогнозируемые классы.

В 2006 г. исследователи [Yu et al., 2008] применили метод опорных векторов и наивный байесовский классификатор для автоматического определения партийной принадлежности авторов текстов (демократ или республиканец) выступлений в Конгрессе США.

В работе [Biessmann, 2016] авторы проанализировали стенограммы выступлений представителей 5 политических партий в немецком бундестаге 17 и 18 созыва, получив среднюю точность 0.61 при использовании метода логистической регрессии.

Анализ 25 000 текстовых сообщений [Prati et al., 2019] из социальных сетей испанского сегмента сети Интернет и применение методов k -ближайших соседей и случайного леса позволило кластеризовать информационное политическое поле, связанное с выборами 2015 г.

Рекуррентные нейронные сети [Yu et al., 2014] активно используются для классификации политических текстов. Модель долгой краткосрочной памяти (LSTM) применена в работе [Rao et al., 2016] для определения текстов сторонников республиканцев и демократов, а в [Li et al., 2017] целевая двунаправленная сверточно-рекуррентная модель – классификации коротких текстов по идеологическому принципу.

Языковая модель двунаправленного кодировщика архитектуры трансформер (BERT) [Devlin et al., 2019] применена в работах [Crothers et al., 2019; Baly et al., 2020] для выявления ключевых слов, чрезмерно используемых в прошлых политических кампаниях в тексте и предсказания сторонников идеологий.

Применение нейросетевых методов для анализа русскоязычного идеологического поля осложнено рядом факторов. В диссертационном исследовании И.Г. Мелешкина [Мелешкин,

2008] делается вывод о том, что для некоторых политических партий (например, «Справедливая Россия») характерна стадия формирования идеологии, в других же случаях идеология отсутствует, заменяясь образом лидера (В.В. Жириновский в ЛДПР). В статьях [Татаркин, 2014; Абзалбеков, 2019] анализируется разница между фактическими ценностями политических объединений и декларируемыми в программных документах идеологиями.

В работе [Schuhmacher et al., 2014] доказана возможность применения графов для моделирования текстовых документов. Китайскими исследователями [Chen et al., 2017] предложен алгоритм выявления описываемой в тексте политической идеологии на основе анализа графа мнений, содержащего связи между сущностями текста и отношения к ним представителей политических движений.

В статье [Feng et al., 2021] описан метод формирования посредством искусственных нейронных сетей графа политических знаний, представленного в виде словаря, в котором понятия расположены на основании степени смысловой близости, т. е. идеографического словаря, или тезауруса. Тезаурусы современного типа не только точно показывают семантические отношения, но и создают модель национального мировосприятия, поэтому отражают действия центров информационного влияния.

Применение векторных языковых моделей, представляющих собой числовое представление графовых структур, для построения семантической сети слов и выявления тезаурусных отношений рассмотрено в русскоязычных диссертационных исследованиях [Установ, 2018; Алексеевский, 2018; Тихомиров, 2022]. В работе К.В. Лунева [Лунев, 2021] тезаурус создается на основе построенного автоматически графа совстречаемости ключевых слов текстов.

Обладая свойством оценочности, идеологемы по-разному воспринимаются сторонниками и противниками политических движений и религиозных течений. В связи с этим актуальной остается проблема определения авторской позиции (stance detection) – автоматического извлечения реакции (положительной, отрицательной, нейтральной) субъекта на заявление, сделанное актором, относительно целевого объекта [Mohammad et al., 2016]. Частным случаем решения этой задачи является определение тональности текста относительно упоминаемого целевого объекта – целевой анализ настроений.

В работе [Hosseini-Asl et al., 2022] описан алгоритм дообучения моделей архитектуры GPT для объектно-ориентированного сентимент-анализа на английском языке. В [Moloshnikov et al., 2023] исследована применимость модели архитектуры T5, состоящей из слоев кодировщика, аналогичных BERT, и слоев декодировщика GPT, для решения задачи для русскоязычных текстов.

Теоретическая модель интеллектуальной системы идентификации идеологем

Идентификация идеологем является многоэтапным процессом, автоматизация которого требует рассмотрения многих факторов и разработку новых методов и алгоритмов.

Проведенный анализ показал, что в настоящий момент существуют различные подходы к применению ЭВМ для решения задачи классификации идеологических категорий в текстах, однако получившие широкое распространение за рубежом модели машинного обучения, обеспечивающие классификацию политических текстов на уровне документов, не могут быть применены для моделирования русскоязычного информационного поля, что потребует разработки алгоритмов автоматического поиска косвенных упоминаний идеологий.

Имеется возможность адаптации модели таргетного анализа настроений русскоязычных текстов для выявления авторской позиции относительно идеологем, выражающих политические идеологии.

Существующие методы построения тезаурусных отношений могут быть применены для построения справочников политических и религиозных ассоциаций на основе контекстов употребления идеологем в новостных статьях.

Теоретическая модель интеллектуальной информационной системы, входными данными которой является текст новостной публикации, а выходными – его идеологемы с указанием идеологии и авторской позиции по отношению к ней, включает следующие подсистемы:

- модуль предварительной обработки текстовых данных;
- модуль генерации религиозно-политических ассоциаций (идеологем текста);
- модуль целевого анализа настроений;
- модуль классификации идеологий;
- модуль итоговой обработки данных.

Модуль генерации религиозно-политических ассоциаций представляет собой обученную с учителем нейросетевую модель преобразования текстовых последовательностей (text-to-text), осуществляющую построение ассоциативных рядов политических и религиозных терминов по аналогии с решением задачи машинного перевода. В качестве набора обучающих данных могут выступать как размеченные экспертами новостные тексты, так и сгенерированные с помощью ассоциативных словарей и моделей машинного обучения датасеты. В модуле должен быть реализован алгоритм поиска фрагментов текста (потенциальные идеологемы), приведших к генерации термина ассоциативного ряда.

Нейросетевой модуль целевого анализа настроений обеспечивает классификацию авторской позиции по отношению к потенциальным идеологемам, в том числе именам публичных личностей, наименований организаций и государств.

Модуль классификации идеологий должен включать семантическую сеть-тезаурус понятий идеологических категорий, созданный в том числе на основе автоматического анализа контекстов употребления слов и словосочетаний русскоязычных политических текстов, и обеспечивать поиск соответствующей религиозной или политической ассоциации идеологии.

Модуль итоговой обработки данных осуществляет аккумуляцию выходных параметров нейросетевых модулей, осуществляя вывод эмоционально окрашенных ассоциативных стимулов и соответствующих им наименований политических идеологий-рефлексов.

На рис. 1 представлена модель интеллектуальной системы классификации идеологем русскоязычных новостных публикаций.



Рис 1. Модель интеллектуальной системы
 Fig 1. The model of the intelligent system

Представленная теоретическая модель описывает процессы, проводимые экспертами-лингвистами в рамках анализа идеологем. Создание программного комплекса потребует решения ряда технических задач, в том числе разработку алгоритмов формирования обучающих наборов данных, содержащих соответствующую назначениям модулей идеологическую разметку, а также выбор метрик оценки качества моделей машинного обучения.

Заключение

Таким образом, проведенное исследование указывает на существенное внимание к проблеме идентификации идеологем как элементов информационного противоборства, а также возможность применения автоматизированных методов для её решения. Приведено обоснование использования подхода на основе искусственных нейронных сетей с учетом адаптации алгоритмов, применяемых для решения смежных задач.

В процессе исследования поставлена проблема автоматической идентификации идеологем как совокупности проблем преобразования последовательностей, целевого анализа настроений и построения тезаурусных отношений, а также разработана теоретическая модель интеллектуальной системы, обеспечивающей выявление идеологем русскоязычного текста.

Список источников

1. Наследов А.Д. 2004. Математические методы психологического исследования: анализ и интерпретация данных: учеб. пособие. СПб.: Речь, (ГПП Печ. Двор), 389 с.
2. Чудинов А.П. 2012. Политическая лингвистика: учебное пособие. 4-е изд. Москва, Флинта, Наука, 254 с.

Список литературы

- Абзалбеков Б.С. 2019. Особенности ценностно-нормативных ориентаций современных политических партий России. Известия Уральского федерального университета. Сер. 3, Общественные науки, 14, 2 (188): 74–84.
- Алексеевский Д.А. 2018. Методы автоматического выделения тезаурусных отношений на основе словарных толкований: диссертация на соискание ученой степени кандидата филологических наук: 10.02.21. Москва, 142 с.
- Журавлев С.А. 2004. Идеологемы и их актуализация в русском лексикографическом дискурсе: диссертация на соискание ученой степени кандидата филологических наук: 10.02.01. Йошкар-Ола.
- Карамова А.А. 2015. Идеологемы: определение понятия и типология. Современные проблемы науки и образования, 2: 1.
- Лунев К.В. 2021. Теоретико-графовые алгоритмы выявления семантической близости между понятиями на основе анализа наборов ключевых слов взаимосвязанных объектов: диссертация на соискание ученой степени кандидата физико-математических наук: 05.13.17. Москва, 195 с.
- Машечкин И.В., Петровский М.И., Поспелова И.И., Царёв Д.В. 2016. Методы автоматического аннотирования и выделения ключевых слов в задачах обнаружения экстремистской информации в сети Интернет. Современные информационные технологии и ИТ-образование, 12(1): 188–198.
- Мелешкин И.Г. 2008. Идеологии политических партий в условиях трансформации партийной системы современной России: диссертация на соискание ученой степени кандидата политических наук: 23.00.02. Санкт-Петербург, 210 с.
- Радина Н.К., Козлова А.В., Набокова А.А. 2018. Методы изучения политического поля: Алгоритм идентификации контекстуальных идеологем (На примере региональной культурной политики). Политическая наука, 2: 252–270.
- Татаркин В.Е. 2014. Идеология политических партий в условиях реформирования партийной системы современной России. Среднерусский вестник общественных наук, 4 (34): 39–44.
- Тихомиров М.М. 2022. Методы автоматизированного пополнения графов знаний на основе векторных представлений: диссертация на соискание ученой степени кандидата физико-математических наук: 05.13.11. Москва, 119 с.
- Усталов Д.А. 2017. Модели, методы и алгоритмы построения семантической сети слов для задач обработки естественного языка: диссертация на соискание ученой степени кандидата физико-математических наук: 05.13.17. Москва, 129 с.
- Яцко В.А. 2016. Распределение хи-квадрат и взвешивание терминов. Символ науки, 3-3: 105–107.
- Baly R, Da San Martino G., Glass J., Nakov P. 2020. We Can Detect Your Bias: Predicting the Political Ideology of News Articles. In Proceedings of the 2020 Conference on Empirical Methods in Natural

- Language Processing (EMNLP), Association for Computational Linguistics: 4982–4991.
- Biessmann F. 2016. Automating political bias prediction.
- Chen W., Xiao Z., Wang T., Yang B., Yi L. 2017. Opinion-aware Knowledge Graph for Political Ideology Detection. Twenty-Sixth International Joint Conference on Artificial Intelligence. IJCAI-17: 3647–3653. DOI 10.24963/ijcai.2017/510
- Devlin J., Chang M.-W., Lee K., Toutanova K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies. Minneapolis, Minnesota. Association for Computational Linguistics. 1: 4171–4186.
- Crothers, E., Japkowicz, N., Viktor, H. L. 2019. Towards Ethical Content-Based Detection Of Online Influence Campaigns. IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP): 1–6. DOI 10.1109/mlsp.2019.8918842
- Feng S., Chen Z., Zhang W., Li Q., Zheng Q., Chang X., Luo M. 2021. KGAP: Knowledge Graph Augmented Political Perspective Detection in News Media. arXiv (Cornell University). DOI 10.48550/arxiv.2108.03861
- Hosseini-Asl E., Liu W., Xiong C. 2022. A generative language model for few-shot Aspect-Based sentiment analysis. arXiv (Cornell University). DOI 10.48550/arxiv.2204.05356.
- Iyyer M., Enns P., Boyd-Graber J., Resnik P. 2014. Political ideology detection using recursive neural networks. In Proceedings of the 52nd annual meeting of the Association for Computational Linguistics. 1: 1113–1122.
- Li X., Chen W., Wang T., Huang W. 2017. Target-Specific Convolutional Bi-directional LSTM Neural Network for Political Ideology Analysis. Web and Big Data. APWeb-WAIM 2017. Lecture Notes in Computer Science. Springer, Cham. 10367: 64–72.
- Lundberg S. M. et al. 2020. From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1): 56–67.
- Mohammad S., Kiritchenko S., Sobhani P., Zhu X., Cherry C. 2016. Semeval-2016 task 6: Detecting stance in tweets. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016): 31–41.
- Moloshnikov I., Skorokhodov M., Naumov A., Rybka R., Sboev A. 2023. Named Entity-Oriented Sentiment Analysis with text2text Generation Approach. In Proceedings of the International Conference “Dialogue”.
- Petrovskiy M., Tsarev D., Pospelova I. 2017. Pattern based information retrieval approach to discover extremist information on the Internet. In Mining Intelligence and Knowledge Exploration: 5th International Conference, MIKE 2017, Hyderabad, India, December 13–15, 2017. Springer International Publishing. 5: 240–249. DOI 10.1007/978-3-319-71928-3_24.
- Prati R. C., Said-Hung E. 2019. Predicting the ideological orientation during the Spanish 24M elections in Twitter using machine learning. AI & SOCIETY. 34: 589–598. DOI 10.1007/s00146-017-0761-0.
- Rao A., Spasojevic N. 2016. Actionable and political text classification using word embeddings and LSTM.
- Schuhmacher M., Ponzetto S.P. 2014. Knowledge-based graph document modeling. Proceedings of the 7th ACM international conference on Web search and data mining: 543–552.
- Yu B., Kaufmann S., Diermeier D. 2008. Classifying Party Affiliation from Political Speech. Journal of Information Technology & Politics. 5: 33–48. DOI 10.1080/19331680802149608.

References

- Abzalbekov, B.S. 2019. Normative and Value Orientations of Modern Political Parties of Russia. *Tempus et Memoria*, 14, 2 (188): 74–84. (in Russian)
- Alekseevskij D.A. 2018. *Metody avtomaticheskogo vydelenija tezaurusnyh otnoshenij na osnove slovarnyh tolkovaniy: dissertacija na soiskanie uchenoj stepeni kandidata filologicheskikh nauk [Methods of Automatic Allocation of Thesaurus Relations based on Dictionary Interpretations: Dissertation for the Degree of Candidate of Philological Sciences]*: 10.02.21. Moscow, 142 p. (in Russian)
- Zhuravlev S.A. 2004. *Ideologemy i ih aktualizacija v russkom leksikograficheskom diskurse: dissertacija na soiskanie uchenoj stepeni kandidata filologicheskikh nauk [Ideologemes and Their actualization in Russian Lexicographic Discourse: Dissertation for the Degree of Candidate of Philological Sciences]*: 10.02.01. Yoshkar-Ola (in Russian)
- Karamova A.A. 2015. Ideologemes: notion definition and typology. *Modern problems of science and education*, 2: 1. (in Russian)

- Lunev K.V. 2021. Teoretiko-grafovye algoritmy vyjavlenija semanticheskoy blizosti mezhdru ponjatijami na osnove analiza naborov ključevykh slov vzaimosvjazannykh objektov: dissertacija na soiskanie uchenoj stepeni kandidata fiziko-matematicheskikh nauk [Graph-Theoretical Algorithms for Identifying Semantic Proximity Between Concepts Based on the Analysis of Keyword of Interrelated Objects: Dissertation for the Degree of Candidate of Physical and Mathematical Sciences]: 05.13.17. Moscow, 195 p. (in Russian)
- Mashechkin I., Petrovskiy M., Pospelova I., Tsarev D. 2016. Automatic summarization and keywords extraction methods for discovering extremist information on the Internet. *Modern Information Technologies and IT-Education*, 12(1): 188–198. (in Russian)
- Meleshkin I.G. 2008. Ideologii politicheskikh partij v uslovijah transformacii partijnoj sistemy sovremennoj Rossii: dissertacija na soiskanie uchenoj stepeni kandidata politicheskikh nauk [Ideologies of Political Parties in the Context of the Transformation of the Party system in Modern Russia: Dissertation for the Degree of Candidate of Political Sciences]: 23.00.02. Saint Petersburg, 210 p. (in Russian)
- Radina N.K., Kozlova A.V., Nabokova A.A. 2018. Methods of studying the political field: The algorithm of contextual ideologemes identification (on the example of regional cultural policy). *Political science (RU)*, 2: 252–270. (in Russian)
- Tatarkin V.E. 2014. Ideologija politicheskikh partij v uslovijah reformirovanija partijnoj sistemy sovremennoj Rossii [The Ideology of Political Parties in the Context of Reforming the party System in Modern Russia.]. *Central Russian Journal of Social Sciences*, 4 (34): 39–44. (in Russian)
- Tihomirov M.M. 2022. Metody avtomatizirovannogo popolnenija grafov znanij na osnove vektornykh predstavlenij: dissertacija na soiskanie uchenoj stepeni kandidata fiziko-matematicheskikh nauk [Methods of Automated Replenishment of Knowledge Graphs Based on Vector Representations: Dissertation for the Degree of Candidate of Physical and Mathematical Sciences]: 05.13.11. Moscow, 119 p. (in Russian)
- Ustalov D.A. 2017. Modeli, metody i algoritmy postroenija semanticheskoy seti slov dlja zadach obrabotki estestvennogo jazyka: dissertacija na soiskanie uchenoj stepeni kandidata fiziko-matematicheskikh nauk [Models, Methods and Algorithms for Building a Semantic Network of Words for Natural Language Processing Tasks: Dissertation for the Degree of Candidate of Physical and Mathematical Sciences]: 05.13.17. Moscow, 129 p (in Russian)
- Jacko V.A. 2016. Raspredelenie Hi-kvadrat i vzveshivanie terminov [Chi-Squared Distribution and Term Weighting]. *Symbol of science*, 3-3: 105–107. (in Russian)
- Baly R, Da San Martino G., Glass J., Nakov P. 2020. We Can Detect Your Bias: Predicting the Political Ideology of News Articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics: 4982–4991.
- Biessmann F. 2016. Automating political bias prediction.
- Chen W., Xiao Z., Wang T., Yang B., Yi L. 2017. Opinion-aware Knowledge Graph for Political Ideology Detection. *Twenty-Sixth International Joint Conference on Artificial Intelligence. IJCAI-17*: 3647–3653. DOI 10.24963/ijcai.2017/510
- Devlin J., Chang M.-W., Lee K., Toutanova K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies. Minneapolis, Minnesota. Association for Computational Linguistics*. 1: 4171–4186.
- Crothers, E., Japkowicz, N., Viktor, H. L. 2019. Towards Ethical Content-Based Detection Of Online Influence Campaigns. *IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*: 1–6. DOI 10.1109/mlsp.2019.8918842
- Feng S., Chen Z., Zhang W., Li Q., Zheng Q., Chang X., Luo M. 2021. KGAP: Knowledge Graph Augmented Political Perspective Detection in News Media. *arXiv (Cornell University)*. DOI 10.48550/arxiv.2108.03861
- Hosseini-Asl E., Liu W., Xiong C. 2022. A generative language model for few-shot Aspect-Based sentiment analysis. *arXiv (Cornell University)*. DOI 10.48550/arxiv.2204.05356.
- Iyyer M., Enns P., Boyd-Graber J., Resnik P. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd annual meeting of the Association for Computational Linguistics*. 1: 1113–1122.
- Li X., Chen W., Wang T., Huang W. 2017. Target-Specific Convolutional Bi-directional LSTM Neural Network for Political Ideology Analysis. *Web and Big Data. APWeb-WAIM 2017. Lecture Notes in Computer Science. Springer, Cham*. 10367: 64–72.



- Lundberg S. M. et al. 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1): 56–67.
- Mohammad S., Kiritchenko S., Sobhani P., Zhu X., Cherry C. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*: 31–41.
- Moloshnikov I., Skorokhodov M., Naumov A., Rybka R., Sboev A. 2023. Named Entity-Oriented Sentiment Analysis with text2text Generation Approach. In *Proceedings of the International Conference “Dialogue”*.
- Petrovskiy M., Tsarev D., Pospelova I. 2017. Pattern based information retrieval approach to discover extremist information on the Internet. In *Mining Intelligence and Knowledge Exploration: 5th International Conference, MIKE 2017, Hyderabad, India, December 13–15, 2017*. Springer International Publishing. 5: 240–249. DOI 10.1007/978-3-319-71928-3_24.
- Prati R. C., Said-Hung E. 2019. Predicting the ideological orientation during the Spanish 24M elections in Twitter using machine learning. *AI & SOCIETY*. 34: 589–598. DOI 10.1007/s00146-017-0761-0.
- Rao A., Spasojevic N. 2016. Actionable and political text classification using word embeddings and LSTM.
- Schuhmacher M., Ponzetto S.P. 2014. Knowledge-based graph document modeling. *Proceedings of the 7th ACM international conference on Web search and data mining*: 543–552.
- Yu B., Kaufmann S., Diermeier D. 2008. Classifying Party Affiliation from Political Speech. *Journal of Information Technology & Politics*. 5: 33–48. DOI 10.1080/19331680802149608.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.

Conflict of interest: no potential conflict of interest related to this article was reported.

Поступила в редакцию 29.01.2024

Received January 29, 2024

Поступила после рецензирования 26.02.2024

Revised February 26, 2024

Принята к публикации 29.02.2024

Accepted February 29, 2024

ИНФОРМАЦИЯ ОБ АВТОРАХ

INFORMATION ABOUT THE AUTHORS

Секерин Алексей Викторович, аспирант кафедры программного обеспечения и администрирования информационных систем, Курский государственный университет, г. Курск, Россия

Alexei V. Sekerin, Postgraduate Student, Department of Software and Administration of Information Systems, Kursk State University, Kursk, Russia

Кудинов Виталий Алексеевич, доктор педагогических наук, профессор, профессор кафедры программного обеспечения и администрирования информационных систем, Курский государственный университет, г. Курск, Россия

Vitaly A. Kudinov, Doctor of Pedagogy, Professor, Professor of the Department of Software and Information Systems Administration, Kursk State University, Kursk, Russia