

УДК 004:338

ЦИФРОВАЯ ТЕНЬ РОССИЙСКИХ МЕЖДУНАРОДНЫХ МЕГАПРОЕКТОВ СТРОИТЕЛЬСТВА АЭС ЗА РУБЕЖОМ: ОЦЕНКА ТОНАЛЬНОСТИ ВЫСКАЗЫВАНИЙ

Гусева А.И., Кузнецов И.А., Бочкарёв П.В., Смирнов Д.С.

ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ», Москва,
e-mail: aguseva@mephi.ru

Данная статья продолжает цикл исследований цифровой тени российских международных мегапроектов строительства объектов атомной генерации за рубежом (на примере Венгрии и Финляндии). Исследование тональности публикаций электронных СМИ, накапливающихся в цифровой тени, необходимо для определения уровня общественной приемлемости атомной энергетики. Для данного исследования использовался подход, включающий в себя следующие шаги: сбор новостных сообщений, посвященных атомной энергетике, в электронных СМИ каждой из стран; предобработка данных; построение имитационной модели для конкретной страны для вычисления вероятности тональности новостной публикации в электронных СМИ на произвольную тему; определение тональности собранных сообщений и вычисление эмпирической вероятности; сравнительный анализ математической и эмпирической вероятностей для определения внешнего информационного влияния; сегментация публикаций и авторов на выделенные группы. Вероятностная имитационная модель строилась методом Монте-Карло на примере Венгрии. Определение тональности новостных сообщений осуществлялось сверточными нейронными сетями CNN, которые хорошо зарекомендовали себя для решения подобных задач. Для сегментации использовался метод двойной ансамблевой кластеризации с изменяющейся метрикой расстояния. В результате сравнительного анализа математической и полученной эмпирической вероятностей в центральных венгерских электронных СМИ внешнее информационное влияние выявлено не было.

Ключевые слова: мегапроект строительства АЭС, информационно-семантическое поле, цифровая тень, тональность высказываний, имитационная модель, сверточные нейронные сети CNN, метод кластеризации с изменяющейся метрикой расстояний

DIGITAL SHADOW OF RUSSIAN INTERNATIONAL MEGAPROJECTS OF NPP CONSTRUCTION ABROAD: ASSESSMENT OF THE TONE OF UTTERANCES

Guseva A.I., Kuznetsov I.A., Bochkarev P.V., Smirnov D.S.

National Research Nuclear University MEPHI (Moscow Engineering Physics Institute), Moscow,
e-mail: aguseva@mephi.ru

This article continues the cycle of research on the digital shadow of Russian international megaprojects for the construction of nuclear generation facilities abroad (on the example of Hungary and Finland). The study of the tonality of electronic media publications accumulating in the digital shadow is necessary to determine the level of public acceptability of nuclear energy. For this study, an approach was used that includes the following steps: collecting news reports on nuclear energy in the electronic media of each country; data preprocessing; building a simulation model for a specific country to calculate the probability of the tonality of a news publication in electronic media on an arbitrary topic; determining the tonality of collected messages and calculating empirical probability; comparative analysis of mathematical and empirical probabilities to determine external information influence; segmentation of publications and authors into selected groups. A probabilistic simulation model was built using the Monte Carlo method on the example of Hungary. The determination of the tonality of news reports was carried out by CNN convolutional neural networks, which have proven themselves well for solving such problems. For segmentation, the method of double ensemble clustering with a variable distance metric was used. As a result of a comparative analysis of mathematical and empirical probabilities obtained in the central Hungarian electronic media, no external informational influence was revealed.

Keywords: NPP construction megaproject, information and semantic field, digital shadow, utterance tonality, simulation model, CNN, clustering method with variable distance metric

В данной статье мы продолжаем анализ информации, накапливаемой в цифровой тени российских международных мегапроектов строительства объектов атомной генерации за рубежом [1, 2]. Под цифровой тенью обычно понимается информация о ком-то или о чем-то, которая накапливается на интернет-серверах, электронных СМИ, социальных сетях и т.д. [3–5]. В данном случае речь идет о публикациях в электрон-

ных СМИ участников информационно-семантического поля мегапроекта. Сентимент-анализ этой информации необходим для определения уровня общественной приемлемости мегапроекта строительства АЭС за рубежом, который характеризуется признанием населением иностранных государств экономических и экологических преимуществ атомной энергетики независимо от возможного риска. Низкий уровень обще-

ственной приемлемости в ряде случаев может привести к задержке строительства АЭС из-за активных протестов населения, поэтому программы лояльности, сопровождающие мегапроект, призваны увеличить уровень его общественной приемлемости [1, 2].

Первым шагом при определении уровня общественной приемлемости является сегментация пользователей по трем факторам: уровень знания об атомной промышленности, установки потребителя и экономические выгоды [1, 2]. На следующем шаге на основе сегментации выделяются фокус-группы (классы), на третьем шаге к ним применяются различные стратегии воздействия на общественное мнение с целью повышения лояльности аудитории к мегапроекту.

Установки потребителя в фокус-группах можно определить, исследовав три составляющие: активность интернет-пользователей, тональность публикаций в электронных СМИ и тональность высказываний в социальных сетях.

В данном случае рассмотрение ограничивается электронными новостными СМИ стран Венгрии и Финляндии. Проекты строительства АЭС в этих странах находятся на стадии проектирования – в Венгрии подрядчик российской стороны АО «Атомстройэкспорт» завершил подготовительные работы и ожидает завершения лицензирования [6]. В Финляндии пока идет подготовка документации, проводятся подготовительные работы, само строительство начнется в 2023 г. [6].

В настоящее время широко применяется несколько подходов к определению тональности текста (сентимент-анализу) [7–9]. Один подход опирается на методы, использующие правила и словари оценочной лексики [7], другой – на методы машинного обучения [8], третий – на теоретико-графовые модели [9], четвертый – гибридный [10]. Среди традиционных методов машинного обучения можно выделить наивный байесовский классификатор и метод опорных векторов [11]. Также активно применяются методы глубокого обучения, среди которых можно выделить Convolutional Neural Networks (CNN) Recurrent Neural Networks (RNN), Deep Belief Networks (DBN) и другие [12–14]. Наиболее перспективным для определения тональности слабоструктурированных текстов, к которым относятся публикации в электронных СМИ, являются нейронные сети CNN [14].

Целью данного исследования является определение уровня общественной приемлемости атомной энергетики в странах Венгрии и Финляндии на основе оценки тональности публикаций в СМИ участников

информационно-семантического поля российских мегапроектов строительства АЭС за рубежом и последующего формирования мероприятий программы лояльности.

Материалы и методы исследования

В данной статье мы ограничимся рассмотрением результатов исследования, направленных на определение тональности высказываний в электронных СМИ участников информационно-семантического поля мегапроекта.

Предлагаемый нами подход содержит следующие шаги.

1. Сбор новостных сообщений, посвященных атомной энергетике, в электронных СМИ.

2. Предобработка собранных данных.

3. Построение имитационной модели для вычисления вероятности тональности (негативная, нейтральная, позитивная) новостной публикации в электронных СМИ на произвольную тему.

4. Определение тональности высказываний.

5. Сравнительный анализ вероятностей: определение информационного влияния.

6. Сегментация публикаций и авторов публикаций на выделенные группы.

Рассмотрим каждый этап более подробно.

Шаг 1. Для исследования тональности публикаций участников информационно-семантического поля мегапроекта были выбраны ведущие электронные СМИ в Венгрии и Финляндии за 2019–2021 гг.

В Финляндии были выбраны следующие источники новостей:

– «Iltalehti» (Вечерняя газета, <https://www.iltalehti.fi>) – вечерняя финская газета, третья по тиражу газета Финляндии, выходит шесть дней в неделю, с понедельника по субботу. В 2010 г. онлайн-версия «Iltalehti» была самым посещаемым веб-сайтом в Финляндии, его посещали 1 937 156 человек в неделю, в 2015 г. сайт был шестым по популярности веб-сайтом в стране;

– «Iltasanomat» (Вечерние новости, <https://www.is.fi/>) – одна из двух известных в Финляндии вечерних газет, вторая по величине газета в стране. Согласно Национальному медиа-исследованию, проведенному в 2019 г., «Iltasanomat» является крупнейшим цифровым медиа в Финляндии и охватывает около 2,5 млн финнов. Посещаемость газеты – 2 290 000 посетителей в день;

– «Helsingin Sanomat» (Хельсинкские новости, <https://www.hs.fi/>) – крупнейшая газета Финляндии, выходит ежедневно, занимает шестое место по популярности среди финских ресурсов;

– «Aamulehti» (Утренняя газета, <https://www.aamulehti.fi>) – ежедневная газета, выходящая в Тампере. Является третьей по популярности газетой Финляндии, с момента своего создания (1881 г.) отличается антироссийским настроением.

В Венгрии для всестороннего рассмотрения электронных СМИ были выбраны следующие источники:

– «Index» (<https://index.hu/>) – венгерский новостной портал, расположенный в Будапеште, является самым посещаемым новостным веб-сайтом в Венгрии с примерно 1,5 млн посещений в день;

– «24.hu» (<https://24.hu/>) – один из крупнейших и наиболее посещаемых новостных сайтов Венгрии в интернете. Издателем сайта является Центральная медиагруппа Zrt. Портал содержит разнообразный контент, включая свежие новости, аналитические материалы, видео, обзоры, обзоры и интервью;

– «Blikk» (<https://www.blikk.hu/>) – венгерская ежедневная бульварная газета, издаваемая в Будапеште, принадлежащая немецко-швейцарской медиакомпания Ringier-Axel Springer;

– «Portfolio» (<https://www.portfolio.hu/>) – финансовая онлайн-газета в Венгрии с числом пользователей в один миллион в месяц. В 2018 г. Portfolio Group вошла в десятку самых читаемых новостных сайтов и 15 самых посещаемых сайтов в Венгрии.

По каждому источнику были собраны две формы, глубина сбора данных – с 2019 по 2021 г., всего 150 недель (табл. 1).

Таблица 1

Формы сбора новостных публикаций

Форма 1 (поля)	Форма 2 (поля)
– Дата новости	– Дата
– Заголовок	– Количество новостей
– Аннотация	
– Основной текст новости	
– Автор	
– Категория	

Шаг 2. Обработка собранных текстовых данных включает дополнительный шаг, связанный с предобработкой таких данных.

В рамках проведенного исследования был сделан следующий набор последовательных шагов для предобработки:

– сбор и загрузка корпусов текста – определяется перечень источников, выполняется загрузка в оперативную память;

– токенизация – превращение текстового документа в набор токенов, где под токеном чаще всего понимается отдельное слово или несколько символов или целые фразы;

– использование тезауруса для формирования связей – для конкретных предметных областей формируется словарь терминов, тезаурус, который представляет собой систему связей этих терминов;

– удаление стоп-слов – удаляются слова с минимальной важностью, к которым относятся предлоги, союзы, местоимения и т.д.;

– удаление символов пунктуации и унификация текста – знаки пунктуации, переноса, пробелов или иных специальных символов могут быть удалены;

– стемминг и/или лемматизация – данная операция направлена на отбрасывание изменяемых частей слова (стемминг) или приведение каждого слова к нормальной форме (лемматизация);

– представление текста в машиночитаемом виде – использование различных подходов для формирования векторного (матричного) представления текстового документа.

Шаг 3. Для оценки вероятности тональности новостных публикаций была разработана вероятностная имитационная модель на основе метода Монте-Карло [15, 16]. Предварительный статистический анализ для построения модели проводился по данным, собранным из четырех источников венгерских электронных новостных СМИ (<https://index.hu/>, <https://24.hu/>, <https://www.blikk.hu/>, <https://www.portfolio.hu/>). На рис. 1 представлена динамика совокупного количества информационных публикаций и его медиана за период 2019–2021 гг., 150 недель.

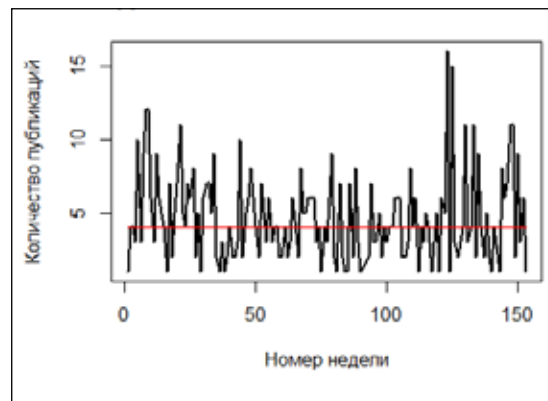


Рис. 1. Динамика совокупного количества публикаций в СМИ по теме «Атомэнергетика» за 2019–2021 гг., 150 недель

На рис. 1 видно, что медиана составляет меньше трети от максимального количества публикаций за неделю. На основании этого, а также текстового анализа заголовков, аннотаций и самих статей новостей была выдвинута гипотеза, что новостная интенсивность реализуется в трех режимах: «событийный», «шумовой» и «тихий».

«Событийный» режим предполагает, что существует некое событие, которое вызывает однотипный отклик у всех новостных источников (тональность такого отклика преимущественно одинакова). К «шумовому» режиму относятся такие интервалы, где появление новостной информации носит спорадический характер и не поддерживается всеми источниками. «Тихая» неделя – общая интенсивность новостных сообщений не превышает одного сообщения в неделю.

Для объективности необходимо ввести еще один тип – «нулевая» неделя, когда нет вообще никаких сообщений на заданную тему. Такой режим преобладает в Финляндии. Например, для одного из анализируемых источников, «Aamulehti», за 150 недель было обнаружено только 15 публикаций по атомной энергетике.

Разделение недель на событийные и шумовые осуществлялось на основании фильтров:

– суммарная новостная интенсивность за неделю превосходит медианную (определенную за 150 недель);

– все исследуемые новостные источники имеют публикации в эту неделю.

Рассмотрим, как менялась интегральная тональность новостных сообщений. Под интегральной тональностью будем понимать сумму тональностей всех сообщений за неделю (обозначение тональности: -1 – негативная, 0 – нейтральная, 1 – позитивная).

На рис. 2, а, б, представлены распределения интегральной тональности для недель, отмеченных как «событийные» и «шумовые».

Для недель этих режимов были рассчитаны частоты новостей разной тональности и математические вероятности этих тональностей (табл. 2).

Статистический анализ событийных недель с преимущественно позитивной и негативной повесткой приведен в табл. 3.

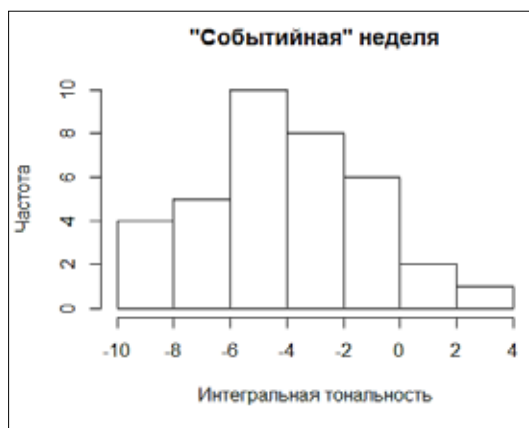


Рис. 2, а. Интегральная тональность «событийной» недели

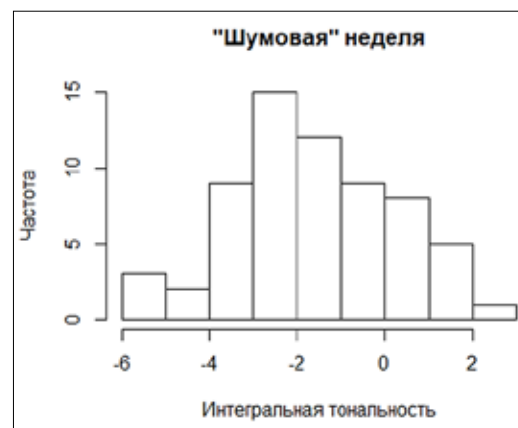


Рис. 2, б. Интегральная тональность «шумовой» недели

Таблица 2

Частоты и вероятности новостей разной тональности

Тональность	«Событийная» неделя		«Шумовая» неделя		«Тихая» неделя	
	Частота	Вероятность	Частота	Вероятность	Частота	Вероятность
Негативная	168	0,7119	119	0,592	20	0,6061
Нейтральная	34	0,1441	35	0,1741	9	0,2727
Позитивная	34	0,1441	47	0,2338	4	0,1212

Таблица 3

Частоты и вероятности новостей «позитивных» и «негативных» недель

Тональность	«Позитивная» неделя		«Негативная» неделя	
	Частота	Вероятность	Частота	Вероятность
Негативная	5	0,2941	163	0,7443
Нейтральная	1	0,05882	33	0,1507
Позитивная	11	0,6471	23	0,105

Эти оценки легли в основу имитационной вероятностной модели, разработанной с помощью метода Монте-Карло. Для каждой недели, в соответствии с ее типом, для каждого моделируемого источника (рассматривалось 10 источников), определялось количество новостных сообщений. Для этого использовались треугольные функции распределения, наиболее часто используемые для решения подобных задач [16].

Таким образом, для каждой недели создавался свой набор сообщений. Их тональность определялась в соответствии с частотами, определенными в ходе статистического анализа. Вероятность негативной тональности сообщения – 0,641, нейтральной – 0,163, позитивной – 0,196. Математическая вероятность необходима для того, чтобы в дальнейшем определить наличие негативного внешнего информационного воздействия в цифровой тени по отношению к мероприятиям программы лояльности.

Шаг 4. Для оценки тональности текста был реализован алгоритм на базе нейронной сети CNN, демонстрирующий высокие результаты на схожих задачах в других предметных областях [14]. Для построения модели использовался вручную размеченный набор данных новостных лент, который составил 50% от общей выборки. Для оставшихся 50% выборки разметка тональности выполнялась с помощью нейронной сети. Реализованная нейронная сеть состояла из четырех скрытых слоев, где процесс обучения включал 10 эпох, реализующих классификацию объекта к одному из трех классов: положительный, нейтральный и отрицательный.

Шаг 5. Сегментация публикаций проводилась на основе полученной ранее экспертной оценки [2], что существует как минимум пять сегментов: позитивной тональности, нейтральной тональности и три сегмента разной степени негативной тональности.

В качестве метода, применяемого для кластеризации новостных публикаций, был использован метод двойной кластеризации с изменяющейся метрикой расстояний. Двойная кластеризация выполнялась в два

шага. На первом шаге проводилась ансамблевая кластеризация по самой новости для публикаций. Каждой публикации после анализа текста присваивается один из пяти классов.

На втором шаге для каждого автора анализировалось количество написанных новостных публикаций (т.е. сколько он написал статей 1-го, 2-го класса и т.д.) и количество написанных им негативных, позитивных и нейтральных сообщений. После этого по этим данным снова проводилась ансамблевая кластеризация с изменяющейся метрикой расстояния и присваивается новый класс уже самому автору.

Результаты исследования и их обсуждение

В табл. 4 приведены полученные результаты оценки тональности публикаций в Финляндии по теме «Ydinvoimala».

Как видно из таблицы, тема атомной энергетики очень мало занимает финское общество. Больше всего новостей, посвященных этой тематике, опубликовано в Helsingin Sanomat. Но даже в ней количество публикаций не превышает 0,1% от общего объема.

Как правило, публикации с негативной тональностью посвящены современным проблемам Фукусимы и Чернобыля, внештатной ситуации на Тайшанской АЭС (Китай), временному закрытию атомной электростанции Словении как меры предосторожности из-за землетрясения в Хорватии и т.д.

Публикации с позитивной тональностью обсуждают достоинства атомной энергетики как безуглеродной, т.е. «зеленой», энергетики. Доминируют утверждения, что ядерная энергия необходима как часть соглашения по климату. Например, единственная положительная публикация в газете «Aamulehti» посвящена обсуждению цен на электроэнергию, которые в 2021 г. очень высокие, но в ближайшее время могут быть снижены за счет введения законопроекта, касающегося строящегося атомного энергоблока.

В табл. 5 приведены результаты оценки тональности новостных публикаций в Венгрии по теме «Atomerőmű».

Таблица 4

Тональность новостных публикаций в Финляндии, посвященных атомной энергетике

Электронные СМИ	Всего новостей (2019–2021 гг.)	Посвящено атомной энергетике	Негативная тональность	Нейтральная тональность	Позитивная тональность
Iltalehti	131016	15	6	6	3
Ilta-Sanomat	166802	32	15	9	8
Helsingin Sanomat	126577	145	57	53	35
Aamulehti	67279	12	8	3	1

Таблица 5

Тональность публикаций в Венгрии, посвященных атомной энергетике

Электронные СМИ	Всего новостей (2019–2021 гг.)	Посвящено атомной энергетике	Негативная тональность	Нейтральная тональность	Позитивная тональность
Index	168468	185	103	38	44
24.hu	131847	176	122	23	31
Blikk	179846	109	82	17	10
Portfolio	61892	199	58	68	73

Таблица 6

Вероятность тональности новостных сообщений (Венгрия)

Тональность новостного сообщения	Вероятность математическая	Эмпирическая вероятность (Венгрия)
Негативная	0,528	0,546
Нейтральная	0,227	0,218
Позитивная	0,243	0,236

Как видно из табл. 5, в Венгрии тема атомной энергетики является более обсуждаемой, чем в Финляндии, количество тематических публикаций составляет 0,3% от общего количества.

К обсуждаемым в Венгрии темам с негативной тональностью добавляются новые, связанные с потеплением воды в Дунае («АЭС Пакш: пошли слухи о потеплении воды Дуная», «Мы вошли в критическую границу: Дунай буквально кипит, объявлена тревога по АЭС Пакш» и т.д.), неисправностями на АЭС Пакш («Авария произошла на одном из блоков АЭС Пакш», «Производительность АЭС Пакш снизилась из-за аварии»), задержками с лицензированием («Пакш 2: выдача очередной лицензии задерживается»), террористические угрозы АЭС («Террористы хотели взорвать атомную электростанцию в Брюсселе», «Это может стать серьезной проблемой безопасности французских атомных электростанций»).

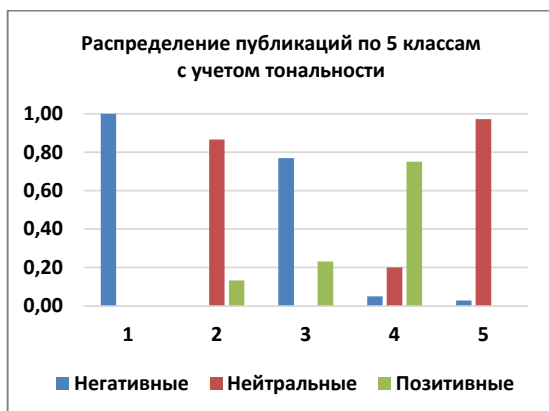
Среди публикаций позитивной тональности, помимо уже упомянутых ранее, появляются публикации, связанные с повышенными мерами безопасности на Пакш-II и экономическими выгодами от реализации мегапроекта («Даже бизнес централизованного теплоснабжения в Пакше принес Lőrinc Mészáros полмиллиарда...», «С осени в университете будет начато обучение строительству АЭС...», «Оказывается, в ближайшие годы в Пакше сможет работать очень много людей...» и т.д.). Результаты исследования показывают, что такой читаемый венгерский источник, как «Portfolio»,

публикует 37% новостей об атомной энергетике позитивной тональности и только 29% – негативной.

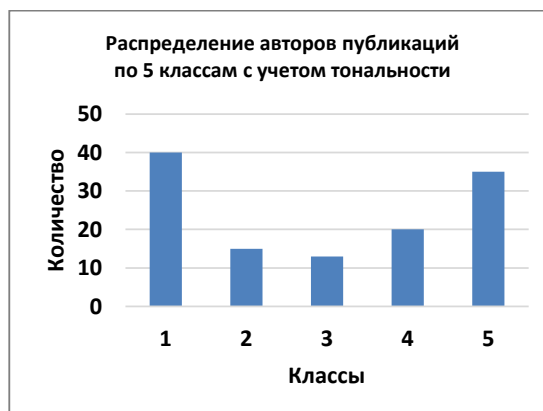
В табл. 6 приведены результаты сравнительного анализа математической вероятности, вычисленной с помощью имитационного моделирования, и эмпирической, полученной в ходе сбора новостных публикаций СМИ в Венгрии. Как видно из таблицы, значения вероятностей очень близки, что говорит об отсутствии какого-либо значительного внешнего информационного воздействия.

На рис. 3 представлены результаты кластеризации самих публикаций новостных электронных СМИ и авторов в Финляндии. «Класс 1» сформирован из публикаций исключительно негативной тональности, «Класс 2» содержит нейтральные публикации с незначительной долей позитивной тональности (15%), «Класс 3» включает негативные публикации с незначительной долей позитивной тональности (20%), «Класс 4» – публикации с основной массой позитивной тональности (75%), незначительной долей нейтральной (20%) и негативной (5%) тональности. «Класс 5» содержит публикации, в подавляющем большинстве нейтральные (95%) с незначительной долей негативных.

Кластеризация авторов по пяти классам показывает, что наиболее объемными являются «Класс 1» и «Класс 5». Количество авторов, входящих в эти классы, в два раза больше, чем в «Классе 4» с позитивной тональностью публикаций.

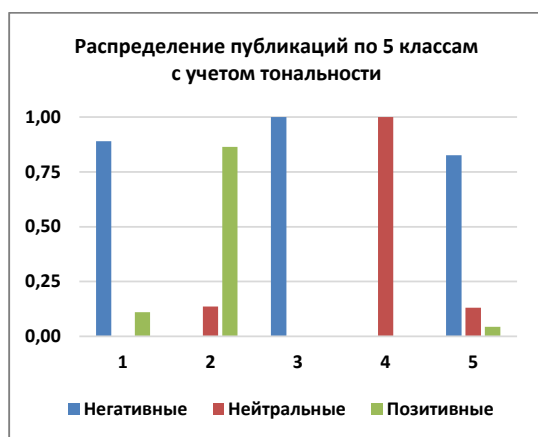


А) Результаты кластеризации публикаций

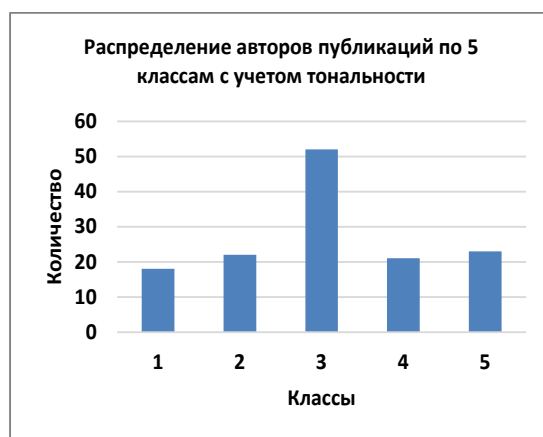


Б) Результаты кластеризации авторов

Рис. 3. Распределение публикаций и авторов по пяти классам с учетом тональности на примере Финляндии (2019–2021 гг.)



А) Результаты кластеризации публикаций



Б) Результаты кластеризации авторов

Рис. 4. Распределение публикаций и авторов по пяти классам с учетом тональности на примере Венгрии (2019–2021 гг.)

На рис. 4 представлены результаты кластеризации авторов венгерских новостных электронных СМИ. Среди выделенных классов «Класс 1» содержит негативные публикации с незначительной долей позитивной тональности (10%); «Класс 2» – позитивные публикации с незначительной долей нейтральной тональности (10%); «Класс 3» включает в себя публикации исключительно негативной тональности; «Класс 4» – публикации исключительно нейтральной тональности, а «Класс 5» имеет смешанный характер и содержит публикации с основной массой негативной тональности (85%), незначительной долей нейтральной (10%) и позитивной (5%) тональности. Кластеризация

авторов по пяти классам показывает, что наиболее объемным являются «Класс 3». Количество авторов, входящих в этот класс исключительно негативной тональности, в два с половиной раза больше, чем в «Классе 2» с положительной тональностью публикаций.

Как мы видим, выделенные классы в Венгрии и Финляндии несколько различаются, что говорит о естественных различиях в установках потребителей по отношению к атомной энергетике в разных странах. Но на основе выделенного сходства можно обобщить полученные результаты следующим образом для обеих стран:

– «Класс А» – в подавляющем большинстве позитивные публикации (75–90%),

с небольшой долей нейтральных и, возможно, негативных.

– «Класс В» – нейтральные публикации, возможно с небольшой долей позитивных высказываний (15%).

– «Класс С» – исключительно публикации с негативной тональностью.

– «Класс D» – негативные публикации с незначительной долей положительной тональности (10–20%).

– «Класс E» – имеет смешанный характер. Это публикации нейтрально-негативные либо негативно-нейтральные с небольшой долей позитивных.

Кластеризация авторов публикаций по этим пяти классам в дальнейшем дает возможность определить стратегию влияния на аудиторию с целью повышения ее лояльности к мегапроектам строительства российских АЭС за рубежом.

Заключение

Обобщая полученные результаты, можно сделать следующие выводы.

Тема атомной энергетики не является популярной в Венгрии и Финляндии, количество новостей, посвященных этой теме за три последних года, составляет не более 0,3–0,1% от общего количества новостных публикаций.

Среди новостных публикаций, посвященных атомной энергетике, большинство имеет отрицательную тональность. Сравнительный анализ математической вероятности, вычисленной с помощью имитационного моделирования, и эмпирической, полученной в ходе сбора новостных публикаций СМИ в Венгрии, показывает, что значения вероятностей очень близки. Это свидетельствует об отсутствии какого-либо значительного внешнего информационно-го воздействия.

Кластеризация публикаций по пяти классам с учетом тональности приводит к выделению класса с публикациями исключительно негативной тональности, двух классов публикаций с преобладающей положительной тональностью и преобладающей нейтральной тональностью, класса с негативной тональностью и незначительной долей положительной тональности (10–20%), класса публикаций смешанного характера (нейтрально-негативные, либо негативно-нейтральные).

Кластеризация авторов публикаций показывает, что объем кластеров с негативными публикациями значительно превышает все остальные кластеры.

Работа поддержана грантом РФФИ № 20-010-00708\21.

Список литературы

1. Ковтун Д.А., Коптелов М.В., Гусева А.И. Управление информационными рисками с помощью информационно-семантического поля в международных проектах атомной энергетики // Современные наукоемкие технологии. 2019. № 11. С. 66–71.
2. Guseva A.I., Matrosova E., Tikhomirova A., Kovtun D. Assessment of the Public Acceptance of the Nuclear Power Plant Construction Plan on the Territory of Foreign Country. Procedia Computer Science. 2021. Vol. 190. P. 301–311.
3. Bergs T., Gierlings S., Auerbach T., Klink A., Schrakneppera D., Augspurgera T. The concept of the Digital Twin and Digital Shadow in Manufacturing. Procedia CIRP 101. 2021. P. 81–84.
4. Ткачева А.В. Цифровой след и цифровая тень потребителя: сущность, формирование и использование // Приоритетные векторы развития промышленности и сельского хозяйства: материалы III Международной научно-практической конференции. Магеевка, 2020. С. 279–285.
5. Бояркина Л.А., Бояркин В.В. Цифровой след и цифровая тень как производные персональных данных // Сборники конференций НИЦ Социосфера. 2016. № 62. С. 78–81.
6. Матросова Е.В., Тихомирова А.Н., Киреев В.С., Гусева А.И. Оценка уровня знаний об атомной промышленности в Белоруссии, Венгрии, Турции и Финляндии // Фундаментальные исследования. 2021. № 12 (1). С. 180–185.
7. Майорова Е.В. О сентимент-анализе и перспективах его применения // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. Реферативный журнал. 2020. № 4. С. 78–86.
8. Гималетдинова Г.К., Довтаева Э.Х. Сентимент-анализ читательского комментария: автоматизированная VS ручная обработка текста // Ученые записки Казанского университета. Серия: Гуманитарные науки. 2021. Т. 163. № 1. С. 65–80.
9. Куницына Н.Н., Метель Ю.А. Применение метода интеллектуального текстового анализа в оценке уровня удовлетворенности клиентов коммерческих банков // Вестник Саратовского государственного социально-экономического университета. 2018. № 2 (71). С. 149–155.
10. Андросов А.Ю., Бородащенко А.Ю., Леонидова К.С. Алгоритм определения тональности публикаций СМИ к должностным лицам государственных органов // Известия Тульского государственного университета. Технические науки. 2020. № 2. С. 47–53.
11. Золина Е.В., Гамова Н.А. Наивный классификатор Байеса для решения задачи сентимент-анализа тестов // Шаг в науку. 2019. № 4. С. 140–142.
12. Романов А.С., Васильева М.И., Куртукова А.В., Мещеряков Р.В. Анализ тональности текста с использованием методов машинного обучения // R. Piotrowski's Readings in Language Engineering and Applied Linguistics. Proceedings. Ser. "CEUR-WS Workshop Proceedings" St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Herzen State Pedagogical University of Russia, 2018. P. 86–95.
13. Черкасов А.Н., Туркин Е.А. Выбор оптимальной архитектуры искусственной нейронной сети для задачи классификации текстов // Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. 2021. № 1 (276). С. 62–66.
14. Kuznetsov I.A., Koptelov M.V., Kovtun D.A. and Guseva A.I. A method for reducing the impact of information risks at numerous stages of a megaproject life cycle based on a semantic information field. Procedia Computer Science. 2021. Vol. 190. P. 500–507.
15. Гильванова Г.А. Анализ риска инновационного проекта методом имитационного моделирования (метод Монте-Карло) // Science Time. 2015. № 12 (24). С. 157–161.
16. Кузикова Д.С., Якушев А.А. Метод Монте-Карло в экономических исследованиях // Тенденции развития науки и образования. 2018. № 34–3. С. 41–45.