

doi: 10.51639/2713-0576_2025_5_2_20

Научная статья

УДК 338: 001.895

ГРНТИ 06.54.31

ВАК 5.2.3

Классификация новостных материалов по географическому признаку и выявление фейковых новостей: современные подходы и методы

Олеся Николаевна Панамарева ^{1*}, Михаил Владимирович Шульженко ²,
Дмитрий Александрович Сухарев ³

^{1, 2} *Военный инновационный технополис «ЭРА»,
Анапа, Россия, *era_otd1@mil.ru*

³ *Войсковая часть 55060, Москва, Россия, 55060-406@mil.ru*

Аннотация

В условиях ведения гибридных войн информация является ключевым ресурсом, использование которого может быть направлено как во благо, так и во вред обществу. Национальная безопасность государства, включающая экономическую, технологическую и другие виды безопасности, а следовательно, и устойчивость всех хозяйствующих агентов находятся в зоне высокого риска. Обладание своевременной и достоверной информацией о различных процессах, объектах, научных и технологических достижениях, инновациях является основой обеспечения национальной безопасности и устойчивого развития общества. При всевозрастающем санкционном давлении на Россию разработка отечественных ИТ-инструментов для детерминирования истинной и ложной информации имеет высокий уровень актуальности. В статье раскрываются современные подходы и методы для классификации новостных материалов по географическому признаку и выявления недостоверных новостей. Анализируется роль географического контекста в повышении точности анализа новостей, а также эффективность различных методов, таких как геокодирование, извлечение именованных сущностей, использование географических информационных систем, проверка фактов, лингвистический анализ и машинное обучение. Особое внимание уделяется интеграции классификации информации по географическому признаку и подходов к определению ее достоверности для борьбы с дезинформацией, как технологического инновационного решения. Показано, что предлагаемый комбинированный подход, учитывающий географический контекст и лингвистические особенности текста, позволит повысить точность выявления фейков. В статье также обсуждаются ограничения существующих методов и обозначаются перспективы дальнейших исследований в данной области.

Ключевые слова: классификация, географический признак, фейковые новости, геокодирование, извлечение именованных сущностей, машинное обучение, инновации, устойчивость, национальная безопасность

Введение

Современное информационное общество характеризуется большим объемом генерируемого и распространяемого контента. Этот огромный поток информации, с одной стороны, предоставляет гражданам возможность быть в курсе происходящих событий, а с другой – создает благоприятную среду для дезинформации и распространения недостоверных (фейковых) новостей и информации [1].

Новости играют ключевую роль в информационных войнах, которые включают использование информации для влияния на общественное мнение, формирования нарративов и достижения стратегических целей. В условиях глобализации и цифровизации на фоне ИТ-трансформаций, нашедших отражение в целом ряде научных работ [2-7], объем информации, распространяемой через Интернет, достигает существенных масштабов. Новостные материалы становятся одним из основных источников информации для широкой аудитории, формируя общественное мнение и определяя повестку дня. Однако не все новости достоверны, и их интерпретация может существенно различаться в зависимости от регионального контекста.

Многие новости имеют региональную или локальную специфику, что значительно усложняет их автоматическую классификацию и категоризацию. Географическая привязка помогает уточнить контекст новости, что особенно важно для событий, которые могут быть интерпретированы по-разному в зависимости от региона. Например, экономические изменения или политические решения могут иметь разнообразные последствия в различных частях мира, а знание и учет их географической привязки позволяет лучше понять их влияние. Последнее снижает риск дезинформации, поскольку пользователи новостных ресурсов получают информацию, адаптированную под их регион. Например, если новость о событии в России содержит ссылки на источники из других стран, это может свидетельствовать о попытке манипулирования информацией.

Исследования, проведенные ЮНЕСКО, подчеркивают важность учета регионального контекста при анализе новостей [8].

Авторы работы [8], акцентируют внимание на том, что дезинформация часто адаптируется под локальные особенности, чтобы вызвать большее доверие у аудитории. Такая информация кажется более релевантной и достоверной для конкретной группы людей, что усиливает ее воздействие. Кроме того, современные исследования Совета Европы показывают, что фейковые новости, адаптированные под локальный контекст, оказывают более сильное влияние на аудиторию, что нашло отражение в исследовании [9], авторы которого представили результат анализа того, как дезинформация распространяется в Центральной Азии и других регионах, преодолевая границы и адаптируясь к местным условиям. Они подчеркивают, что без учета географического контекста становится невозможным полноценный анализ новостей и выявление потенциальной дезинформации.

Классификация новостных материалов по географическому признаку (КНМГП) становится важным инструментом в борьбе с дезинформацией. Она позволяет не только учитывать региональные особенности событий, но и выявлять аномалии в распространении новостей. Например, если новость о событии в одной стране активно распространяется в другом регионе, это может указывать на координированную кампанию по дезинформации. Таким образом, интеграция КНМГП и анализа достоверности новостей открывает новые возможности для повышения точности детерминирования недостоверной информации, для противодействия манипуляциям в области информационной сферы, и как следствие – для снижения и нивелирования рисков, связанных с ведением гибридных войн, ставших объективной реальностью.

Следовательно, для эффективного анализа новостных материалов необходимо решить две взаимосвязанные задачи:

- осуществление КНМГП, необходимой для определения их территориальной принадлежности и контекста;
- выявление недостоверных новостей для минимизации влияния дезинформации на общество.

Особенности классификации новостных материалов по географическому признаку

КНМГП позволяет анализировать новостной контент в его пространственном контексте [10]. Этот процесс включает в себя определение и присвоение географического тега – местоположения – региона, страны, города и др.

Современные подходы к КНМГП включают в себя использование таких методов, как:

- геокодирование;
- извлечение именованных сущностей;
- применение географических информационных систем (ГИС).

Геокодирование – это процесс преобразования текстовых упоминаний географических объектов (например, названий городов, стран, адресов) в географические координаты (широта и долгота).

Алгоритмы геокодирования могут быть основаны на использовании словарей географических названий, статистических методов или гибридных подходов, сочетающих различные техники.

Словарные методы геокодирования предполагают использование баз данных, содержащих информацию о географических названиях и их координатах. При обработке новостного текста алгоритм ищет соответствия между словами в тексте и записями в базе данных, и при нахождении совпадений присваивает новостному материалу соответствующие координаты.

Для преобразования текстовых упоминаний географических объектов в географические координаты используются словарные, статистические и гибридные алгоритмы. При этом основными инструментами являются GeoPy, Nominatim, Google Maps API. GeoPy, Nominatim, Google Maps API. Например, если новость содержит такой текст, как «Вчера в Париже произошла забастовка», то словарный метод геокодирования найдет в базе данных запись для «Парижа» и присвоит новости соответствующие координаты (48.8566°, 2.3522°).

Извлечение именованных сущностей (Named Entity Recognition, NER) – категория задач в области Natural Language Processing (NLP). NER позволяет выделять из текста географические названия (например, города, страны, регионы). Для этого используются как традиционные лингвистические модели, так и современные архитектуры на основе глубокого обучения, такие как BERT, SpaCy или Stanford NLP, достигающие точности до 90 % в извлечении географических сущностей. Например, можно рассмотреть ситуацию, где из текста «В Берлине прошли переговоры между лидерами Германии и другими представителями стран Европы» будут выделены слова «Берлин», «Германия» и «Европа».

Географические информационные системы позволяют интегрировать новостные данные с географическими данными, проводить пространственный анализ и визуализировать результаты. Например, можно использовать карты плотности для визуализации концентрации новостей в определенных регионах, анализ кластеров для выявления групп новостных событий, расположенных вблизи друг друга, и пространственную корреляцию для оценки связи между новостными событиями и различными географическими факторами (например, плотностью населения, уровнем экономического развития, уровнем технологического развития и политической ситуацией).

При решении обозначенных выше задач используются открытые базы данных, такие как GeoNames, Google Geocoding API и OpenStreetMap, которые позволяют сопоставлять текстовые данные с реальными местоположениями.

Для сравнения методов геокодирования информации в таблице 1 приведены их основные характеристики.

КНМГП может улучшить анализ новостей, позволяя исследователям выявлять региональные тенденции, пространственные закономерности и соотносить новостные события с реальными географическими объектами.

Таблица 1 – Сравнительные характеристики методов геокодирования информации

Характеристика метода геокодирования		Наименование метода геокодирования			
		Словарный метод	Статистический метод	Гибридный метод	<i>NER</i>
1	Достоинства	Простота	Учет контекста	Высокая точность	Автоматизация
2	Недостатки	Неоднозначность	Требует много данных	Сложность реализации	Зависит от данных
3	Точность	~ 80 %	~ 85 %	~ 95 %	~ 90 %
4	Инструменты	<i>GeoNames</i>	Машинное обучение (<i>ML</i>)	Словарь + <i>ML</i>	<i>BERT, SpaCy</i>

Так анализ географического распределения новостных сообщений об изменении климата может выявить регионы, особенно уязвимые к его последствиям, а исследование географического распределения информационных сообщений об изменении отношения к ситуации, связанной с проведением Специальной военной операции РФ на Украине, позволит выявить стороны (участников), находящихся в зоне наибольшего риска и влияния возможных угроз. КНМГП тесно связана с проблемой достоверности новостных материалов, эффективность совместного решения данных задач существенно отражается на качестве и оперативности принимаемых решений.

Особенности выявления недостоверных новостей

Проблема достоверности новостных материалов активно исследуется в последние годы. Фальшивые (фейковые) новости представляют собой заведомо ложную информацию, распространяемую с целью манипулирования общественным мнением, нанесения ущерба репутации или достижения политических/экономических целей [11].

Современные подходы к выявлению фейков включают в себя:

- 1) анализ источника и автора;
- 2) проверку фактов (*fact-checking*);
- 3) лингвистический анализ;
- 4) методы машинного обучения (МО) и технологии искусственного интеллекта (ИИ).

Ключевые характеристики методов детерминирования фейковых новостей приведены в таблице 2.

1. Анализ источника и автора является одним из первых шагов в процессе выявления фейков. Он предполагает оценку репутации и надежности новостного источника, проверку наличия предвзятости и связей с заинтересованными сторонами, а также анализ стиля написания и языковых характеристик автора.

Методы оценки репутации и надежности новостных источников включают в себя анализ трафика на сайт, проверку WHOIS-информации (данных о владельце домена), анализ связей с другими организациями и лицами, а также оценку предыдущих публикаций источника.

Таблица 2 – Сравнительные характеристики методов определения фейковых новостей

Характеристика метода определения фейковых новостей		Наименование метода определения фейковых новостей			
		Анализ источника и автора	<i>Fact-checking</i>	Лингвистический анализ	<i>ML</i>
1	Достоинства	Простой, быстрый	Точность, эксперты	Автоматика, объем текста	Автоматизация, высокая точность
2	Недостатки	Субъективный, сложные связи	Трудоемкий, сложно автоматизировать	Нет контекста, адаптация	Зависит от данных
3	Точность	~ 80 %	~ 85 %	~ 95 %	~ 90 %
4	Инструменты	<i>WHOIS</i>	<i>PolitiFact</i>	<i>NLP</i>	<i>SVM, RNN, LSTM</i>

Анализ стиля написания и языковых характеристик автора может выявить признаки, указывающие на возможную предвзятость или недостоверность информации. Например, использование эмоционально окрашенной лексики, грамматические ошибки и несоответствие стилю авторитетных изданий могут быть признаками фейка. В последнее время все большую популярность приобретает привлечение пользователей для оценки репутации и надежности новостных источников на основе их личного опыта и знаний.

2. Проверка фактов (*fact-checking*) – это процесс сопоставления информации, представленной в новостном материале, с другими источниками, экспертной оценкой и доступными данными. Целью проверки фактов является установление достоверности информации и выявление возможных неточностей, искажений или фальсификаций [12].

Существует ряд инструментов и баз данных для проверки фактов, таких как Snopes, PolitiFact, FactCheck.org. Эти инструменты предоставляют пользователям возможность проверять утверждения, представленные в новостных материалах, и получать информацию об их достоверности.

В последнее время разрабатываются автоматизированные методы проверки фактов на основе баз знаний и анализа текстового сходства. Эти методы позволяют автоматически сопоставлять утверждения, представленные в новостном материале, с информацией, содержащейся в базах знаний и других источниках, и определять степень их соответствия. Основными проблемами автоматической проверки фактов являются сложность выявления контекста, необходимость экспертной оценки и ограниченность доступных баз знаний.

3. Лингвистический анализ предполагает использование методов обработки естественного языка (NLP) для анализа текста новостей. Его целью является выявление языковых признаков, характерных для фейковых новостей (например, использование эмоционально окрашенной лексики, кликбейтные заголовки, грамматические ошибки и низкая читаемость).

Методы NLP, используемые для лингвистического анализа, включают в себя:

- анализ тональности текста;
- выявление ключевых слов и фраз;
- анализ структуры предложений и синтаксических связей;
- оценку читаемости текста.

Результаты лингвистического анализа могут быть использованы для создания автоматических информационно-аналитических систем, которые используют языковые признаки для определения степени достоверности новостного материала.

4. Методы МО и технологии ИИ. Алгоритмы МО могут быть обучены на большом количестве размеченных данных (т.е. новостных материалов, для которых известна их

достоверность) для выявления закономерностей и признаков, характерных для фейковых новостей.

Существует множество моделей МО, используемых для выявления фейков таких как логистическая регрессия, метод опорных векторов (Support Vector Machines, SVM), дерево принятия решений (Decision Tree), нейронные сети и др.

В последнее время все большую популярность приобретают методы глубокого обучения, такие как рекуррентные нейронные сети (RNN), долго-краткосрочная память (LSTM), которые позволяют анализировать текст и метаданные новостей с высокой точностью. Алгоритмы МО, такие как случайный лес (Random Forest), SVM, позволяют автоматически классифицировать новости как достоверные или фейковые [13]. Например, модель МО может быть обучена на наборе новостей с различными характеристиками (источник, автор, заголовок, текст, метаданные) и затем использоваться для классификации новых новостей.

Основными проблемами МО в контексте выявления фейков являются смещение данных (т.е. наличие предвзятости в обучающих данных), необходимость большого количества размеченных данных и интерпретируемость моделей.

Рассмотрим существующие решения определения недостоверных новостей. В качестве примеров использования МО для выявления фейков можно рассмотреть такие инструменты, как FakeBananas и FakeBox. Эти платформы демонстрируют различные подходы к автоматизации анализа новостного контента.

FakeBananas применяет метод сравнения позиций статей с авторитетными источниками, тогда как FakeBox фокусируется на анализе заголовков, текстового содержания и метаданных новостей. Оба сервиса используют алгоритмы МО, что позволяет достигать высокой точности, хотя их эффективность может варьироваться в зависимости от качества входных данных и контекста.

Сервис FakeBananas [14] представляет собой инструмент, разработанный с применением методов МО для идентификации недостоверных новостей. Он был создан в 2017 г. командой участников хакатона HackMIT. Основная цель FakeBananas – помочь пользователям определить, насколько достоверной является новостная информация. Для этого сервис использует метод stance detection, который сравнивает позицию статьи или утверждения с позициями других источников. Это позволяет оценить, насколько утверждение согласуется с мнением авторитетных изданий.

FakeBananas принимает на вход ссылку на статью или произвольное утверждение. Затем он ищет тысячи статей, связанных с ключевыми словами, и анализирует их позицию по отношению к данному утверждению.

Модель FakeBananas основана на сочетании нескольких методов (таблица 3).

Таблица 3 – Методы в составе модели *FakeBananas* для определения фейковых новостей

Наименование метода		Описание / пояснение
1	<i>Bag-of-Words</i>	представление текста как набора слов без учета их порядка
2	<i>Google's word2vec</i>	алгоритм для представления слов в виде векторов, учитывающих их семантическое значение
3	<i>TF-IDF</i>	метод оценки важности слов в тексте
4	«Стоп-слова»	исключение часто встречающихся, но малозначимых слов (например, предлогов и союзов)

FakeBananas способен достаточно эффективно определять достоверность новостей, хотя результаты могут варьироваться в зависимости от качества входных данных и контекста. Точность модели по состоянию на сегодня достигает 82 % на тестовых данных.

Хотя FakeBananas предоставляет рекомендации, он не всегда объясняет, почему было принято то или иное решение. Это может затруднить понимание логики работы алгоритма. Кроме того, инструмент фокусируется только на текстовых данных и не учитывает такие факторы, как географический контекст, метаданные или визуальный контент.

FakeBox – это уникальный сервис, созданный компанией Machine Box, предназначенный для автоматического анализа новостного контента с целью оценки его достоверности. Этот инструмент использует передовые методы МО и технологии ИИ, чтобы классифицировать новости как правдивые или недостоверные [15].

В основе работы FakeBox лежит анализ различных аспектов новостной статьи, включая заголовок, текстовое содержание и метаданные, такие как URL. Сервис легко интегрируется в сторонние приложения через REST API, что делает его удобным решением для автоматизации проверки новостей.

FakeBox анализирует различные аспекты статьи, такие как:

- 1) заголовок (оценивается на наличие кликбейтных элементов, эмоционально окрашенной лексики и других признаков недостоверности);
- 2) содержание текста (проверяется информация на предмет языковых особенностей, таких как грамматические ошибки, несоответствие стилю авторитетных изданий и использование преувеличенных утверждений);
- 3) URL статьи: (анализируется доменное имя и другие метаданные для оценки репутации источника).

После завершения анализа сервис оценивает вероятность того, является ли новость фейковой. Чем больше информации предоставлено (например, заголовок, текст и URL), тем выше точность результатов. На тестовых данных FakeBox демонстрирует точность выше 95 %. Однако в реальных условиях эффективность системы может снижаться из-за сложности контекста и ограниченности данных. Но, несмотря на высокую точность, сервис имеет недостатки:

- точность анализа снижается при неполных или некорректных входных данных;
- модель требует регулярного переобучения, так как язык новостей и способы распространения дезинформации постоянно меняются;
- FakeBox фокусируется исключительно на текстовых данных, игнорируя такие важные факторы, как географический контекст или визуальный контент.

Использование КНМГП позволяет связать новость с конкретным регионом или страной, что помогает:

- 1) проверить достоверность источников (если новость о событии в одной стране ссылается на источники из другой, это может быть признаком дезинформации);
- 2) обнаруживать аномалии в распространении новостей (если новость о локальном событии (например, землетрясении в Японии) активно распространяется в других регионах, это может указывать на попытку манипуляции общественным мнением);
- 3) контекстуализировать информацию (географический контекст помогает интерпретировать новости в зависимости от региональных особенностей).

В целом географический контекст может быть полезен для выявления фейковых новостей. Например, если новостное сообщение утверждает, что в определенном городе произошел инцидент, который не подтверждается никакими другими источниками и противоречит общему контексту ситуации в этом городе, то это может служить признаком того, что сообщение является фейковым. А если новость о событии, произошедшем в определенном регионе, распространяется преимущественно в других регионах, то это может свидетельствовать о том, что новость является частью скоординированной кампании по дезинформации.

Анализ связи между географическими факторами (такими как уровень экономического развития, политическая ситуация, доступность информации) и распространением дезинформации может помочь выявить регионы, наиболее подверженные

влиянию фейковых новостей, и разработать адресные стратегии противодействия и проактивные меры.

Интеграция методов классификации новостных материалов по географическому признаку и современных подходов к выявлению фейковых новостей

Исходя из сказанного выше, интеграция КНМГП и современных подходов к выявлению фейков является перспективным направлением в области обеспечения информационной безопасности. Такой комплексный подход позволит:

- проверять соответствие новостного сообщения географическому контексту, и, следовательно, выявлять недостоверные новости, которые используют географические аномалии (например, ссылки на источники из других стран) или противоречат региональному контексту;
- повысить точность анализа и более эффективно бороться с дезинформацией;
- получить более глубокое понимание географических закономерностей распространения фейковых новостей, что особенно актуально в условиях повышения интенсивности ведения гибридных войн, разрушающих экономическую мощь государств, регионов;
- выявлять регионы, наиболее подверженные влиянию дезинформации;
- разрабатывать адресные стратегии противодействия фейковым новостям и проактивные меры.

Перспектива интеграции КНМГП и современных подходов к выявлению фейков обусловлена необходимостью создания специализированных датасетов и ИТ-ресурсов, разработки алгоритмов, учитывающих географический контекст и лингвистические особенности приемов дезинформации, а также решения этических и социальных вопросов, связанных с использованием географической информации для выявления фейков.

Для успешной интеграции КНМГП и современных подходов к выявлению фейковых новостей необходимо создание больших объемов, размеченных данных, содержащих информацию о географической привязке новостных материалов и их достоверности. Эти данные могут быть использованы для обучения алгоритмов МО и оценки их эффективности. Разработка алгоритмов, учитывающих географический контекст и лингвистические особенности фейковых новостей, требует привлечения специалистов в области географии, лингвистики, обработки естественного языка и МО. Следовательно, целесообразно использовать комбинацию таких инструментов, как географический контекст (анализ местоположения, упоминаемого в новостях), лингвистический анализ (выявление языковых признаков фейковых новостей) и МО (использование алгоритмов для автоматической классификации новостей).

Заключение

В данной работе рассмотрены современные подходы и методы для классификации новостных материалов по географическому признаку и выявления фейковых новостей. Основные выводы:

1. Всевозрастающее санкционное давление на Россию, необходимость формирования и сохранения технологического (в первую очередь в области ИТ-технологий – драйвера экономического развития на инновационной основе) суверенитета разработка и внедрение отечественных ИТ-инструментов для детерминирования истинной и ложной информации имеет высокий уровень актуальности
2. КНМГП важна для анализа контекста и повышения точности анализа информации.

3. Современные методы, такие как NER, геокодирование и МО в сочетании с технологиями ИИ, показывают высокую эффективность, позволяя анализировать новостной контент в его пространственном контексте.

4. Интеграция КНМГП и современных подходов к выявлению фейковых новостей открывает новые возможности для борьбы с дезинформацией в целях обеспечения устойчивости, экономической и национальной безопасности в целом.

Использование КНМГП и современных подходов к выявлению фейковых новостей для выявления дезинформации поднимает ряд этических вопросов, связанных с защитой частной жизни и обеспечением свободы выражения мнения. Таким образом, остаются открытыми вопросы, связанные с качеством данных, этическими аспектами и необходимостью создания универсальных моделей. Эти и связанные с ними направления требуют дальнейших исследований.

Конфликт интересов

Авторы статьи заявляют, что на момент подачи статьи в редакцию, у них нет возможного конфликта интересов с третьими лицами.

Список источников

1. Уордл К., Дерахшан Х. Информационный беспорядок: к междисциплинарной основе для исследований и разработки политики. – URL: <http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf> // (дата обращения: 10.01.2025). – Текст: электронный.
2. ИТ в 2023-м году: заглянем в будущее // Тренды на Рынке ИТ (it-world.ru). – URL: <https://www.it-world.ru/it-news/reviews/189340.html?ysclid=lfhywgaxrf43071307> // (дата обращения: 21.03.2025). – Текст: электронный.
3. Морозов А.В., Панамарев Г.Е., Гусеница Я.Н. Состояние и перспективы развития современной науки в области информационно-телекоммуникационных технологий в военном инновационном технополисе «ЭРА» // Состояние и перспективы развития современной науки по направлению «ИТ-технологии». Сборник трудов II Всероссийской научно-технической конференции. Анапа: Военный инновационный технополис «ЭРА». – 2023. – С. 7-18.
4. Научные и технологические тренды: 2020-2040 гг.: перспективы научно-технического развития: реферат доклада организации НАТО по науке и технологиям. 2020 // Science & Technology Trends 2020-2040: Exploring the S&T Edge / NATO Science & Technology Organization. Brussels. – 2020. – 153 p.
5. Панамарев Г.Е., Галенко И.С. Способ формирования модуля реестра перспективных технологий и инновационных проектов // Состояние и перспективы развития современной науки по направлению «АСУ, информационно-телекоммуникационные системы»: Сборник статей III Всероссийской научно-технической конференции. Анапа. – 2021. – С. 46-53.
6. Тренды и технологии 2030 (ict.moscow). – URL: <https://ict.moscow/research/trendy-i-tehnologii-2030/?ysclid=lfhze77tjf741756762> // (дата обращения: 21.03.2025). – Текст: электронный.
7. McKinsey: 10 главных технологических трендов до 2030 года для компаний и рынков – CDO2DAY. – URL: <https://cdo2day.ru/cifrovoj-analiz/mckinsey-10-glavnyh-tehnologicheskikh-trendov-do-2030-goda-dlja-kompanij-i-rynkov/?ysclid=lfhz7t0rer229276765> // (дата обращения: 21.03.2025). – Текст: электронный.
8. Муратова Н., Тошпулатова Н., Алимова Г. Fake news: дезинформация в медиа: Пособие для студентов направлений журналистики и массовых коммуникаций. – URL:

<https://unesdoc.unesco.org/ark:/48223/pf0000374919> // (дата обращения: 12.01.2025). – Текст: электронный.

9. Вестербай С.Д., Джураев Ш., Маразис А. Медиа потребление и дезинформация в центральной Азии: количественная и качественная оценка в контексте геополитики. – URL: <https://encouncil.org/wp-content/uploads/2023/05/2023-03-ENC-Report-RUS-V5.pdf> // (дата обращения: 10.01.2025). – Текст: электронный.

10. Zandbergen P.A. Geocoding // ESRI Press, 2008. – URL: https://www.researchgate.net/publication/26129516_Geocoding_School_and_Student's_Home_Addresses_Zandbergen_Responds // (дата обращения: 15.01.2025). – Текст: электронный.

11. Lazer D. M. J., et al. The science of fake news [Электронный ресурс] // Science. – 2018. – Vol. 359, № 6380. – P. 1094–1096. – URL: https://www.researchgate.net/publication/323650280_The_science_of_fake_news // (дата обращения: 16.01.2025). – Текст: электронный.

12. Graves L. Deciding What's True: The Promise and Challenge of Fact-Checking. – New York: Columbia University Press, 2016. – 320 p.

13. Мальчиц В.С., Гетман А.Н. Обработка данных для машинного обучения и применение метода опорных векторов для реализации классификатора новостей // Вестник Амурского государственного университета. Серия: Естественные и экономические науки. 2019. – № 87. – URL: <https://cyberleninka.ru/article/n/obrabotka-dannyh-dlya-mashinnogo-obucheniya-i-primenenie-metoda-opornyh-vektorov-dlya-realizatsii-klassifikatora-novostey> // (дата обращения: 20.01.2025). – Текст: электронный.

14. Основной источник проекта FakeBananas. – URL: <https://github.cosm/likeaj6/FakeBananas> // (дата обращения: 06.02.2025). – Текст: электронный.

15. Основной источник проекта FakeBox. – URL: <https://machinebox.io/docs/fakebox> // (дата обращения: 11.02.2025). – Текст: электронный.

Classification of news materials by geography and identification of fake news: modern approaches and methods

Olesya Nikolaevna Panamareva ^{1*}, Mikhail Vladimirovich Shulzhenko ²,
Sukharev Dmitry Aleksandrovich ³

^{1*,2} Innovativ Technopolis «ERA», Anapa, Russia, *era_otd1@mil.ru

³ Military Unit 55060, Moscow, Russia, 55060-406@mil.ru

Abstract

In the context of hybrid warfare, information is a key resource, the use of which can be directed both to the benefit and to the detriment of society. National security of the state, including economic, technological and other types of security, and, consequently, the stability of all economic agents are in the high-risk zone. Having timely and reliable information about various processes, objects, scientific and technological achievements, innovations is the basis for ensuring national security and sustainable development of society. With the ever-increasing sanctions pressure on Russia, the development of domestic IT tools for determining true and false information is highly relevant. The article reveals modern approaches and methods for classifying news materials by geographic feature and identifying unreliable news. The role of geographic context in increasing the accuracy of news analysis, as well as the effectiveness of various methods, such as geocoding, named entity extraction, the use of geographic information systems, fact checking, linguistic analysis and machine learning, are analyzed. Particular attention is paid to the integration of classification of information by geographic feature and approaches to determining its reliability to combat disinformation, as a technological innovative solution. It is shown that the proposed

combined approach, taking into account the geographic context and linguistic features of the text, will improve the accuracy of identifying fakes. The article also discusses the limitations of existing methods and outlines prospects for further research in this area.

Keywords: classification, geographic feature, fake news, geocoding, named entity extraction, machine learning, innovation, sustainability, national security