

МРНТИ 19.01.07

ИНСТРУКЦИЯ ДЛЯ ЭКСПЕРТНОГО АНАЛИЗА И КЛАССИФИКАЦИИ ДОСТОВЕРНОСТИ ПОДОЗРИТЕЛЬНЫХ ПУБЛИКАЦИЙ НОВОСТНОГО ХАРАКТЕРА

М. Атанаева¹, Т. Булдыбаев², У. Оспанова³, И. Акоева⁴

¹магистр государственного управления, заместитель президента

²заместитель директора Департамента по работе со СМИ и населением

³магистр менеджмента, менеджер Департамента статистики и сопровождения информационных систем

⁴менеджер Департамента аналитики, мониторинга и оценки образования

^{1,2,3,4}АО «Информационно-аналитический центр» МОН РК,

Казахстан, Астана, email: Miraim.Atanayeva@iac.kz

Данная статья подготовлена в рамках реализации ПЦФ № BR05236839 по теме «Разработка информационных технологий и систем для стимулирования устойчивого развития личности как одна из основ развития цифрового Казахстана»

В настоящей статье предложен подход к проведению экспертного анализа корпуса публицистических текстов. По результатам библиографического и социологического опроса приведены показатели оценки вероятной степени достоверности публикации новостного характера. Приведен анализ факторов, который учитывался при формулировании показателей экспертной оценки для классификации достоверности подозрительных публикаций новостного характера. Сформулированы вопросы для экспертного анализа достоверности публикации новостного характера.

Ключевые слова: оценка, экспертно-аналитическая рамка, средства массовой информации, публикация, достоверность

Средства массовой информации (СМИ) играют центральную роль в информировании общественности и соответственно в формировании мировоззрения социума, а также общественного поведения.

Мнение индивида по определенным вопросам формируется под воздействием двух основных источников информации. Первый источник – это окружение человека, его семья, друзья, коллеги, знакомые и другие люди, находящиеся в непосредственном контакте с ним. Второй – общедоступные источники информации, такие как книги, журналы, интернет, СМИ и другие. Если общедоступные источники информации справочного характера, такие как энциклопедии, учебники, книги и другие, помогают сформировать фундаментальные знания, то СМИ, без всякого сомнения, являются основным источником новостей, которые ориентируют общество в отношении того, что происходит в мире, стране в настоящее время. Именно СМИ являются ключевыми игроками в определении повестки дня и фокусируют общественный интерес на конкретных темах, что действует как ограничитель диапазона аргументов и точек зрения, которые выступают информационной основой общественных дискуссий. Стремительное развитие информационных технологий лишь усилило процесс подготовки и распространения информации, в том числе подозрительной. При этом новые каналы распространения информации позволяют более адресно и «по назначению» информировать определенные слои социума и «призывать» к определенным действиям.

По данным глобального исследования настроений Edelman Trust Barometer (Барометр доверия, почти семь из десяти охваченных исследованием респондентов 28 стран мира обеспокоены тем, что фейковые новости и дезинформация используется сегодня как оружие [1]. В этой связи одним из главных вызовов начала 21-го века является умение оперативно распознавать и нейтрализовывать информацию.

На сегодняшний день разработка подходов с целью определения информативных признаков достоверности в текстовой информации является весьма актуальной. В связи с этим одной из главных

задач исследования является выявление и классификация информативных признаков достоверности публикации, для возможного использования автоматизированного подхода в оценке влияния открытых текстовых информационных источников на социум с минимальной вовлеченностью экспертов. Однако необходимо осознавать, что для автоматизации процесса оценки текстовой информации с использованием методов машинного обучения, на первоначальном этапе необходим корпус текстов с экспертной классификацией и глубокой детальной лингвистической разметкой.

Казахстанское экспертное сообщество не остается в стороне и пытается решить поставленную задачу начиная с разработки методологии и формирования базы данных для машинного обучения – проанализированного и детально размеченного корпуса статей. В ходе исследования источником текстовой информации новостного характера выступили публикации новостных интернет порталов. Почему были выбраны именно интернет порталы?

Во-первых, итоги социологического исследования по оценке влияния открытых информационных источников на социум, показали, что для большинства респондентов из Казахстана (59,8%) основным источником получения информации новостного характера являются следующие источники: социальные сети и блогеры (30,3%), а также новостные интернет-сайты (29,5%). [2]

Во-вторых, одним из важнейших трендов в современной коммуникативистике является резкое повышение влияния онлайн и электронных СМИ. Результаты исследования Агентства GlobalWebIndex, свидетельствуют о преобладании цифровых источников информации, над традиционными СМИ, под которыми понимается просмотр телевидения. [3].

Для активной работы с корпусом текстов необходима разметка корпуса — главная характеристика корпуса; она отличает корпус от простых коллекций (или «библиотек») текстов. Под разметкой понимается процесс или результат приписывания текстам и их компонентам специальных меток. Существует два вида разметки, применяющиеся для работы с корпусами текстов: лингвистическая и нелингвистическая/экстралингвистическая (метаразметка). Под метаразметкой понимаются данные, содержащие такие сведения как даты, авторы, места и др. Структура текстов корпуса зависит от поставленных исследованием задач — это может быть анализ отдельных слов, словосочетаний, групп слов, предложений, абзацев, и т.д. [4]

До начала организации процесса классификации необходимо проработать методы извлечения информации, подходящие для дальнейшей обработки и интерпретации с прицелом на дальнейшее кодирование экспертных суждений.

По результатам исследования было установлено, что подходы и критерии к определению достоверности публикаций едины. [5] Вместе с тем большинство методик определения достоверности сведений, информации, данных, а также публикаций основаны в основном на экспертных оценках.

Для эффективной организации экспертного анализа и получения объективных результатов первоначальной задачей является разработка экспертно-аналитической рамки анализа и инструкции как элементов управления процессом и регламентации деятельности.

Обобщенный алгоритм организации сбора экспертных оценок состоит из следующих шагов:

Шаг 1: подготовительный этап для экспертного анализа.

Шаг 2: сбор экспертных оценок.

Шаг 3: обработка и анализ экспертных оценок [6].

Организация процесса экспертного анализа достоверности публикаций новостного характера состояла из следующих этапов:

1. постановка целей и задач экспертного анализа;
2. определение целевой группы экспертов;
3. разработка рамки экспертного анализа;
4. подготовка инструкции экспертного анализа;
5. мониторинг и контроль экспертного анализа;
6. сбор, обработка и анализ экспертных оценок.

Применительно к экспертным суждениям необходимо осознавать важность формулирования информативных признаков, так как в результате субъективных экспертных оценок можно получить подтверждение либо опровержение сформулированной гипотезе. Адекватность и правдивость полученной по итогам экспертных оценок модели в основном и зависит от четко сформулированных показателей информативных признаков: вычленение отдельных факторов, их идентификация и

конечная интерпретация – задачи, решение которых обуславливают успешность результата. При этом одна из главных целей при сборе экспертных оценок – это получить количественную средневзвешенную оценку (математическую модель) по итогам качественных экспертных оценок для дальнейшей автоматизации процесса. [7]

При формулировании показателей экспертной оценки при классификации достоверности подозрительных публикаций новостного характера учитывались следующие факторы:

- сложность в разделении критериев оценки (отношение к публикации от отношения к теме публикации, тональность публикации от тональности темы публикации и т.д.);
- сложность в формулировании и, как следствие, размытость, нечеткость некоторых показателей, что усугубляет субъективизм экспертных оценок и сложность их дальнейшей конвергенции;
- одновременное наличие в публикации диаметрально противоположных критериев (позитивно-негативная тональность);
- необходимость определения информативного признака отдельного показателя (генерализация в публикации может быть, как положительной, так и отрицательной);
- воздействие на экспертную оценку комментариев к публикациям, а именно: тема публикации, как и сама публикация, могут быть нейтральными, но вызвать резкий резонанс в обществе, который можно проследить по негативным высказываниям к статье;
- сложность в вычленении признака при использовании автором статьи противоположных техник манипулирования общественным мнением (например, тема положительна, предоставление материала нейтрально, но в самой статье какой-либо аспект противопоставляется как негативный);
- ввиду яркой эмоциональности комментариев и, казалось бы, легкости определения их окрашенности сложность в определении тональности (например, если в статье написано о случившемся преступлении, а комментарии - это негодование в отношении случившегося факта) и др.

Учитывая, что главной целью исследования является разработка методики определения степеней вероятной достоверности отдельной публикации, целью экспертного анализа являлась апробация выбранных информативных признаков достоверности. По результатам библиографического анализа и итогов социологического исследования были определены следующие, наиболее приемлемые показатели оценки вероятной степени достоверности публикации новостного характера:

- наличие ссылки в публикации на компетентный источник;
- ссылка на первоисточник информации;
- наличие автора публикации;
- рейтинг/репутация издания (сайты новостных и информационных агентств, электронных СМИ, иных аналогичных сайтов), на котором размещена публикация;
- освещение одного и того же события различными изданиями: кроссчекинг, наличие дискурса с другими публикациями;
- соответствие заголовка публикации содержанию;
- наличие проверяемых фактов в статье.

Выбор данных индикаторов достоверности обусловлен, с одной стороны, возможностью последующей автоматизации процесса их проверки; с другой, основан на мнении экспертов (по итогам экспертного анализа).

При этом необходимо учитывать, что для построения в дальнейшем математической модели необходимо будет учесть иерархию признаков: при прочих равных условиях, какой фактор будет являться наиболее весомым, а какой не будет иметь сильного воздействия. Одним из подходов, возможность применения которого высоко вероятна, может служить абстрактность и целостность признака. К примеру, если признак «ссылка на первоисточник» или «наличие проверяемых фактов в статье» - это целостные показатели информативного признака достоверности, которые легко подсчитать и вывести в количественном выражении, то показатели «наличие манипуляторных техник» или «степень политизированности публикации» - это уже иное свойство качества признака. Их количественное выражение будет напрямую зависеть от субъективной оценки эксперта.

В этой связи важной задачей являлось определение надежности кодировщика, т.е. эксперта. Строгий подход к контент-анализу для обеспечения максимальной надежности требует работы двух или более кодировщиков, по крайней мере, для определенного образца контента (называемого суб-

выборкой надежности). Даже когда основной кодировщик выполняет большую часть работы, важна проверка тестовой выборки вторым или иногда третьим кодировщиком, для исключения или максимального снижения уровня субъективности или риска возникновения субъективности. В связи с этим следующий этап проведения работ заключался в качественном формировании пула экспертов. В данном случае во главу угла ставится компетентность экспертов – квалификация эксперта в определенной области знаний. Для определения компетентности анализировалась сфера деятельности экспертов, уровень и широта знаний и пониманий о происходящих в обществе тенденциях, понимание современных трендов общества, видение перспектив развития современного мира и т.д. Определить наличие либо отсутствие указанных качеств и критериев возможно только обследовав окружение и практическую жизнедеятельность эксперта, опросив его окружение (коллег по работе), изучив его трудовые и социальные достижения, интервьюировав его.

Исходя из целей и задач исследования, к экспертному анализу были привлечены практикующие журналисты, специалисты в области политологии, государственного управления, аналитики в различных сферах. Главными критериями при отборе экспертов выступили область профессиональной деятельности, специальность, профессиональные навыки и опыт работы.

Работа экспертов заключалась в проведении оценки публикаций по факторам достоверности. Для объективной разметки корпуса, экспертами был применен метод триангуляции. [8] То есть эксперты заочно были объединены в 3 группы по 3 человека. Каждой группе предоставлялись одни и те же публикации в одном и том же количестве. Таким образом, каждая группа оценивала одни и те же публикации. По результатам проводилось сопоставление результатов и выявление конвергенции между кодировщиками.

Второй этап заключался в формализации правил и формировании словарей по информативным признакам достоверности, которые определялись лингвистическими методами. Экспертам предоставлялся размеченный корпус по классам и параметрам, а также представительный и сбалансированный корпус текстов.

На третьем этапе проводилась работа по формированию рамки экспертного анализа – непосредственно самого инструмента, который стандартизировал экспертные оценки. Для организации всеобъемлющего процесса необходимо руководствоваться следующими основными критериями:

- структурированность и последовательность признаков;
- единообразие в отображении и формулировании;
- ясность изложения и отсутствие двусмысленности в понимании и т.д.

Для нивелирования негативных последствий, следующий этап заключался в разработке инструкции экспертного анализа.

Практика показывает, что инструкция для экспертного анализа должна основываться на принципах простоты в понимании и использовании всеми экспертами, а также конкретности в изложении. Учитывая индивидуальную характер работы каждого эксперта и возможность наличия несогласованности либо субъективной интерпретации признаков, целесообразно провести общий вводный инструктаж по каждому признаку.

Экспертный анализ достоверности публикаций был проведен путем разметки 5 210 текстов из 5 источников информационных интернет-порталов. В инструкции была дана формулировка достоверности: «достоверная публикация – публикация, правдиво, полно и правильно отображающая реальность».

С целью того, чтобы эксперты определили текст достоверным, были сформулированы вопросы, приведенные во вставке. Таким образом, читая данные вопросы, эксперты должны были оценить текст по трем шкалам: достоверный, не достоверный, трудно определяемый (в данном случае была проведена процедура снижения размерности, т.е. выделение наиболее информативных признаков).

1) Считаете ли Вы, что информация, указанная в публикации, является достоверной?

- A. Да
- B. Нет

2) Имеются ли в тексте ссылка на компетентный источник (государственные органы РК, их официальные представители и интернет-ресурсы; национальные компании их официальные представители и интернет-ресурсы, аналитические отчеты и исследования авторитетных

международных организаций, статистические базы данных, профессиональные экспертные площадки, авторитетные научные организации)?

- A. Да
- B. Нет

3) Соответствует ли заголовок содержанию текста?

- A. Да
- B. Нет

4) Указано ли имя автора статьи? (исследования доказывают, что в целом, среди анонимных публикаций, количество недостоверных существенно выше, чем в публикациях, в которых указано имя автора).

- A. Да
- B. Нет

5) Содержит ли текст факты, достоверность которых в дальнейшем можно проверить? (цифры, данные исследований, описание событий со ссылкой на источник, который можно легко найти и проверить точность информации и др.).

- A. Да
- B. Нет

6) Позволяет ли репутация ресурса считать информацию, приведенную в тексте, надежной?

- A. Да
- B. Нет

7) Наличие проверяемых фактов в статье

- A. Да
- B. Нет

По результатам оценки достоверности проводился анализ конвергенции между кодировщиками. Для данной оценки возможно использовать существующие онлайн-калькуляторы (к примеру, <https://nlp-ml.io/jg/software/ira/>). По итогам вышеперечисленных процедур определялась достоверность текста.

В ходе проведения анализа экспертами проводился систематический мониторинг оценки. Главной целью мониторинга было определение единого понимания между экспертами вопросов инструкции.

Обработка полученного результата демонстрирует степень достоверности проведенной экспертизы публикаций. Так, высокая корреляция между экспертными оценками свидетельствует о:

- высокой степени надежности экспертного мнения;
- правильного и прямого формулирования показателей информативного признака;
- достаточности инструктивного материала, сопровождающего рамку экспертного анализа;
- полноты и ясности сформулированных разъяснений.

По результатам экспертного анализа достоверности публикаций экспертами также были оценены тексты по критериям объективности/предвзятости, потенциальной резонансности, тональности, политизированности и наличию манипулятивных техник. Полученные результаты оценки позволяют классифицировать публикации следующим образом:

Классификация 1	Классификация 2
Достоверный	Недостоверный
Достоверный, но предвзятый	Не достоверный, но объективный
Достоверный, но политизированный	Недостоверный и политизированный
Достоверный, потенциально резонансный	Недостоверный и потенциально резонансный
Достоверный, с негативной тональностью	Недостоверный, с негативной тональностью Недостоверный, с нейтральной тональностью Недостоверный, с позитивной

	тональностью
Достоверный, но с наличием манипулятивных техник	Недостоверный и с наличием манипулятивных техник

Таким образом информация, приведенная в таблице, позволяет классифицировать достоверность публикации по различным критериям. Использование инструкции предоставляет возможность формирования единого мнения о целях и задачах экспертного анализа. Для этого инструкция должна быть лаконична, четко структурирована и содержать конкретные и детализированные вопросы.

Список литературы

- 1 Глобальное исследование «Барометр доверия Эдельман 2018» [Электронный ресурс] // URL: http://cms.edelman.com/sites/default/files/2018-02/2018_Edelman_Trust_Barometer_Global_Report_FEB.pdf (дата обращения: 29.07.2018)
- 2 Результаты социологического опроса по оценке влияния открытых информационных источников (электронных СМИ) на социум / АО «Информационно-аналитический центр» МОН РК; - Астана, 2018
- 3 Digital vs. Traditional Media Consumption. Analyzing time devoted to online and traditional forms of media at a global level, as well as by age and across countries. INSIGHT REPORT SUMMARY | Q1. 2017
- 4 Демская-Кульчицкая О.М., Семеренко В.Р., Ющенко Р.А. Методы автоматической разметки текстов национального корпуса языка. *Компьютерная математика*. 2005, № 1
- 5 Rubin, V., Conroy, N., and Chen, Y., Towards News Verification: Deception Detection Methods for News Discourse. 2015
- 6 Ерошкин Сергей Юрьевич, Поляков Владимир Валентинович Анализ подходов к оценке экспертной информации при прогнозировании инновационных решений. Научная статья по кибернетике. *Научные труды: Институт народнохозяйственного прогнозирования РАН. ВАК*, 2005. URL: <https://cyberleninka.ru/article/n/analiz-podhodov-k-otsenke-ekspertnoy-informatsii-pri-prognozirovanii-innovatsionnyh-resheniy> (дата обращения: 27.09.2018)
- 7 Анохин А.Н. Методы экспертных оценок. Учебное пособие. – Обнинск: ИАТЭ, 1996. – 148с. URL: https://www.researchgate.net/publication/265964844_Metody_ekspertnyh_ocenok (дата обращения: 27.09.2018)
- 8 Кошарная Г.Б., Кошарный В.П. Триангуляция как способ обеспечения валидности результатов эмпирического исследования. Известия ВУЗов. Поволжский регион. *Общественные науки*. 2016. №2 (38). URL: <https://cyberleninka.ru/article/n/triangulyatsiya-kak-sposob-obespecheniya-validnosti-rezultatov-empiricheskogo-issledovaniya> (дата обращения: 05.09.2018)

ЖАҢАЛЫҚТАР СИПАТЫНДАҒЫ КҮДІКТІ БАСЫЛЫМДАРДЫҢ СЕНІМДІЛІГІН САРАПТАМАЛЫҚ ТАЛДАУ ЖӘНЕ ЖІКТЕУ

М.К. Атанаева¹, Т. К. Булдыбаев², У.А. Оспанова³, И.Г. Акоева⁴

¹ мемлекеттік басқару магистрі, Президент орынбасары

² БАҚ және жұртшылықпен жұмыс жөніндегі департамент директорының орынбасары

³ менеджмент магистрі, Статистика және ақпараттық жүйелерді сақтау бөлімінің менеджері

⁴ Сараптама, мониторинг және білім беруді бағалау бөлімінің менеджері

^{1,2,3,4} ҚР БҒМ «Ақпараттық-талдау орталығы» АҚ,

Қазақстан, Астана, email: Miraim.Atanayeva@iac.kz

Осы мақалада публицистикалық мәтіндер түрлеріне сараптамалық талдауын өткізу үшін тәсілдеме ұсынылады. Библиографиялық пен социологиялық сауалнаманың нәтижелері бойынша жаңалықтарға жататын жарияланымның анықтығының ықтимал дәрежедегі бағалау көрсеткіштері келтірілді. Жаңалықтарға жататын күдікті жарияланымдардың анықтығын жүйелеу үшін сараптамалық бағалау көрсеткіштерін тиянақтауда ескеріліп отырған факторларының талдауы келтіріліп отыр. Жаңалықтарға жататын жарияланымдардың анықтығының сараптамалық талдауы үшін сауалдар тұжырымдалды.

Түйін сөздер: бағалау, сараптамалық-талдау шеңбері, бұқаралық құралдар, жарияланым, анықтық

**INSTRUCTIONS FOR EXPERT ANALYSIS AND CLASSIFICATION
OF THE RELIABILITY OF SUSPICIOUS NEWS**

M.K. Atanaeva¹, T. K. Buldybaev², U.A. Ospanov³, I.G. Akoeva⁴

¹ Master of public administration, Deputy president

² deputy director of the Department for work with the media and the public

³ Master of management, manager of the Department of statistics and maintenance of information systems

⁴ manager of the Department of analytics, monitoring and evaluation of education

^{1,2,3,4} JSC “Information-Analytical Center” MES RK,
Kazakhstan, Astana, email: Miraim.Atanayeva@iac.kz

This article proposes an approach to expert analysis of the corpus of journalistic texts. From bibliographic and sociological survey, we determined the indicators for assessing the probable reliability of a news media text. We conducted the analysis of the factors, which was taken into account when formulating expert assessment indicators to classify the reliability of doubtful news articles. We drew up a questionnaire and a framework for expert analysis of the reliability of a news media text.

Key words: *assessment, expert-analytical framework, mass media, media-text, reliability*

**ТАРИХ
ЭКОНОМИКА
ҚҰҚЫҚ**

**ИСТОРИЯ
ЭКОНОМИКА
ПРАВО**

**HISTORY
ECONOMICS
LAW**