

ПОИСК ЗНАЧИМОЙ ИНФОРМАЦИИ В УСЛОВИЯХ ИНФОРМАЦИОННОГО ШУМА: ПРОБЛЕМЫ ОТБОРА, ФИЛЬТРАЦИИ И ОЦЕНКИ ДАННЫХ

Ахмедов М.А., студент группы СГНЗ-61Б

Московский государственный технический университет им. Н.Э. Баумана

*Научный руководитель: Бочарников И. В., доктор политических наук,
профессор кафедры «Информационная аналитика и политические технологии»*

Аннотация: Статья посвящена проблеме поиска значимой информации в условиях информационного шума. Рассматриваются трудности отбора, фильтрации и оценки данных в информационно-аналитической работе. Особое внимание уделяется проявлениям информационного шума: дублированию сведений, противоречивости источников, смешению фактов и оценок, а также алгоритмической персонализации информационного потока. Подчеркивается необходимость проверки релевантности, достоверности, актуальности и качества информации перед ее использованием в аналитических выводах.

Ключевые слова: информационный шум, значимая информация, информационно-аналитическая деятельность, отбор данных, фильтрация информации, качество данных, достоверность, релевантность, оценка информации.

Современная цифровая среда характеризуется постоянным ростом объемов информации, в том числе в социальной сфере. Цифровые платформы, средства массовой информации, государственные информационные системы, социальные сети, открытые данные, экспертные публикации и пользовательские сообщения формируют непрерывный информационный поток. На первый взгляд, расширение доступа к данным должно облегчать аналитическую работу. Однако на практике избыток сведений нередко приводит к противоположному результату: возрастает сложность поиска действительно значимой информации, увеличивается риск ошибочного отбора источников, усложняется проверка достоверности и снижается прозрачность аналитического вывода.

В научной литературе данная проблема обычно рассматривается через понятие информационного шума. Информационный шум определяется как поток неотфильтрованных, избыточных или слабо структурированных сообщений, которые затрудняют восприятие информации и принятие решений [7]. Его проявлениями могут выступать дублирующиеся сведения, нерелевантные материалы, устаревшие данные, эмоционально окрашенные сообщения, непроверенные факты, противоречивые оценки и информационные повторы [9]. В социальной аналитике информационный шум особенно опасен, поскольку он способен исказить представление о реальном состоянии социальных процессов и создавать

видимость значимости тех сведений, которые фактически не имеют достаточной аналитической ценности.

Проблема поиска значимой информации связана не только с количеством данных, но и с характером самой информационно-аналитической деятельности. Она предполагает не простое накопление сведений, а их преобразование в новое знание, пригодное для понимания ситуации и принятия решений [6]. В этом процессе можно выделить два взаимосвязанных уровня: информационный и аналитический. Информационный уровень связан с получением, предварительной обработкой, систематизацией и отсеиванием несущественных сведений. Аналитический уровень предполагает более глубокое исследование материала, выявление закономерностей, связей, тенденций и формирование выводов [6]. Следовательно, информационный шум проявляется уже на информационном уровне, а его последствия переходят на аналитический уровень, когда ошибочно отобранные или недостаточно проверенные данные становятся основанием для аналитических заключений.

Первая ключевая проблема состоит в определении самой значимости информации. Значимая информация не равна любой найденной информации. Она должна соответствовать цели исследования, быть связанной с анализируемой социальной проблемой и обладать достаточным качеством для дальнейшей интерпретации [6]. При отсутствии четко сформулированной аналитической задачи практически невозможно отделить полезные сведения от второстепенных. В результате аналитик может собрать большой массив материалов, но не получить содержательной основы для вывода. Такая ситуация особенно характерна для цифровой среды, где поиск часто подменяется накоплением ссылок, сообщений и фрагментов данных.

Вторая проблема связана с различием между релевантностью и аналитической пригодностью информации. Релевантность обычно означает соответствие материала запросу или теме поиска. Однако формально релевантный источник не всегда является полезным для анализа. Материал может совпадать с ключевыми словами, но быть устаревшим, поверхностным, методологически слабым или недостоверным. Поэтому для аналитической работы важна не только релевантность, но и пертинентность, то есть соответствие информации реальной потребности пользователя или исследовательской задаче [8]. В условиях информационного шума это различие имеет принципиальное значение: поисковая выдача может быть обширной, но значимыми окажутся только те сведения, которые помогают раскрыть объект анализа.

Третья проблема заключается в дублировании и тиражировании информации. В цифровой среде один и тот же информационный повод может многократно повторяться в разных источниках, переходить из пер-

вичной публикации в перепечатки, обзоры, комментарии и пользовательские сообщения. В исследованиях информационного шума отмечается, что избыточность и повторяемость сообщений затрудняют выделение содержательно новых данных из общего потока [9]. В медиатекстах шумовой эффект может усиливаться за счет многократного воспроизведения одних и тех же смыслов, эмоциональных акцентов и оценочных конструкций [5]. Поэтому количество публикаций само по себе не может рассматриваться как показатель достоверности или аналитической значимости.

Четвертая проблема связана с противоречивостью данных. Информационный поток часто включает сведения, различающиеся по времени получения, методике сбора, охвату аудитории, целям публикации и способу интерпретации. В результате разные источники могут не совпадать между собой или давать противоположные оценки одного и того же явления. В стандартах качества данных согласованность рассматривается как одна из характеристик, необходимых для оценки пригодности данных к использованию [2]. Для официального статистического учета также значимы сопоставимость, точность и интерпретируемость данных [3]. Следовательно, противоречивость не должна автоматически устраняться путем выбора одного удобного источника. Она требует проверки методики, контекста и границ применимости каждого массива информации.

Пятая проблема состоит в смешении фактов, оценок и интерпретаций. В информационном потоке фактические сведения часто подаются вместе с эмоциональными характеристиками, прогнозами, экспертными мнениями и оценочными суждениями. Для аналитики в социальной сфере это особенно важно, поскольку оценочные материалы могут отражать общественные настроения, но не должны автоматически восприниматься как фактическое описание ситуации. Если аналитик не отделяет факт от интерпретации, то вывод начинает зависеть не от проверенных данных, а от тональности источников [5]. В условиях информационного шума это ведет к усилению субъективности и снижению качества анализа.

Шестая проблема связана с алгоритмической средой поиска. Современный пользователь получает информацию не только через самостоятельный поиск, но и через рекомендательные системы, новостные агрегаторы, поисковую выдачу, персонализированные ленты и системы искусственного интеллекта. Эти механизмы не являются нейтральными. Они ранжируют материалы по собственным техническим и поведенческим принципам, которые не всегда совпадают с критериями научной или аналитической значимости [4]. Поэтому высокая видимость информации в цифровой среде не должна восприниматься как доказатель-

ство ее ценности. Для аналитика важно сохранять критическое отношение к источникам и не подменять аналитический отбор автоматической выдачей.

Решение поднятых проблем начинается с процедуры отбора данных. Отбор представляет собой целенаправленное выделение информации, соответствующей поставленной задаче. В информационно-аналитической деятельности этот этап связан с получением, предварительной обработкой, классифицированием, определением основных зависимостей и отсеиванием несущественных деталей [6]. Поэтому отбор должен строиться на заранее определенных критериях: тематическая связь с проблемой, надежность источника, актуальность, возможность проверки, полнота, методологическая прозрачность и пригодность для анализа. Отбор не может быть полностью механическим, поскольку значимость информации определяется не только внешними признаками источника, но и смысловой связью с исследуемым явлением.

После отбора необходима фильтрация информации. Если отбор отвечает на вопрос, какие сведения потенциально нужны для анализа, то фильтрация позволяет исключить из этого массива лишние, повторяющиеся, нерелевантные и сомнительные данные. В текстовых источниках фильтрация может быть направлена на удаление дублей, отделение фактических сообщений от эмоциональных оценок, исключение рекламных и манипулятивных материалов. В количественных данных она связана с выявлением пропусков, выбросов, несопоставимых показателей и ошибок структуры [10]. Основная трудность фильтрации состоит в необходимости сохранить баланс: избыточно мягкая фильтрация оставляет шум, а чрезмерно жесткая может удалить важные сигналы.

После отбора и фильтрации необходима оценка качества данных. В современных стандартах качество данных раскрывается через совокупность характеристик, среди которых выделяются достоверность, точность, полнота, актуальность, согласованность, доступность, понятность и соответствие цели использования [2]. Эти критерии позволяют перейти от интуитивной оценки источника к более системной проверке его пригодности.

Достоверность информации означает возможность доверять ее содержанию и происхождению. Она зависит от надежности источника, прозрачности методики, возможности проверки и отсутствия признаков искажения [11]. Актуальность показывает, насколько данные соответствуют текущему состоянию изучаемого процесса. Полнота характеризует достаточность сведений для анализа. Точность указывает на степень соответствия данных реальному положению дел. Согласованность позволяет оценить отсутствие необъясненных противоречий между разными элементами информации. В совокупности эти критерии формируют ос-

нову для проверки того, может ли информация использоваться как аналитически значимая.

Особое значение имеет оценка официальных данных. В социальной сфере официальная статистика часто выступает базовым источником анализа, поскольку обладает нормативно закрепленными правилами сбора и обработки. Вместе с тем ее использование также требует оценки качества. В российском подходе к официальному статистическому учету подчеркивается значение достоверности, полноты, научной обоснованности, своевременности и общедоступности статистической информации [1]. Кроме того, современные стандарты качества данных официального статистического учета обращают внимание на такие характеристики, как точность, своевременность, интерпретируемость и сопоставимость [3]. Следовательно, даже официальные данные не должны использоваться без понимания их методики, границ применения и ограничений.

Работа с информационным шумом требует последовательного аналитического цикла. Сначала формулируется проблема и определяется цель поиска. Затем выбираются источники и собираются данные. После этого проводится первичный отбор, фильтрация и очистка массива. Далее данные оцениваются по критериям достоверности, актуальности, полноты, точности и согласованности. Только после этого возможна интерпретация, выявление связей, тенденций и формирование вывода. Нарушение этой последовательности повышает риск того, что аналитическое заключение будет построено на случайных, непроверенных или нерелевантных сведениях.

Поиск значимой информации в условиях информационного шума представляет собой не только техническую, но и методологическую проблему. Главная трудность заключается не в отсутствии доступа к сведениям, а в необходимости отличить аналитически ценные данные от информационного фона. Избыточность источников, дублирование сообщений, противоречивость данных, смешение фактов и оценок, а также влияние алгоритмической среды усложняют работу аналитика и требуют системных процедур проверки.

В заключение следует отметить, что информационный шум является устойчивым признаком современной цифровой среды, в том числе в социальной сфере. Поэтому информационно-аналитическая работа должна быть направлена не на максимальное накопление сведений, а на их осмысленную селекцию, фильтрацию, оценку качества и интерпретацию. Только при соблюдении этих условий информация может быть преобразована в знание, пригодное для понимания социальных процессов и принятия обоснованных управленческих решений.

Литература и источники:

1. Федеральный закон от 29.11.2007 № 282-ФЗ «Об официальном статистическом учете и системе государственной статистики в Российской Федерации».
2. ГОСТ Р 71484.2-2024 (ИСО/МЭК 5259-2:2024). Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 2. Показатели качества данных. М.: Российский институт стандартизации, 2024. 40 с.
3. ГОСТ Р 72297-2025. Государственное управление. Качество данных официального статистического учета. М.: Российский институт стандартизации, 2025. 28 с.
4. Баранова С.В. Релевантность как инструмент управления запросом в условиях трансформации сетевого медиатекста Web 2.0–3.0 // *НОМОТНЕТИКА: Философия. Социология. Право*. 2024. Т. 49. № 3. С. 553–562.
5. Басовская Е.Н. Информационный шум как воздействующий компонент медиатекста // *Вестник Челябинского государственного университета*. 2014. № 7 (336). Филология. Искусствоведение. Вып. 89. С. 6–11.
6. Бочарников И.В. Информационно-аналитическое сопровождение специальных и полицейских операций : учебное пособие. М.: ИНФРА-М, 2026. 267 с.
7. Вертинова А.А., Пашук Н.Р., Макогонова П.В., Кошелева А.И. Оценка влияния информационного шума на принятие решений // *Лидерство и менеджмент*. 2022. Т. 9. № 3. С. 877–890. DOI: 10.18334/lim.9.3.116218.
8. Грекова Л.В. Информационный поиск в информатике и библиотековедении // *Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика*. 2013. № 1-1 (144). Т. 25. С. 95–100.
9. Игнашин А.А., Котлярова В.В. Феномен информационного шума // *Форум молодых ученых*. 2019. № 6 (34). С. 521–525.
10. Копырин А.С., Видищева Е.В. Технологии обработки и очистки данных, выявления и устранения шумов на временном ряду // *Вестник Академии знаний*. 2020. № 4 (39). С. 220–228.
11. Кубраков Д.В. Количественное оценивание достоверности информации в автоматических интеллектуальных системах организаций // *Труды МАИ*. 2025. № 144.
12. Ремарчук В.Н. Информационно-аналитическая деятельность: проблемы и перспективы // *Вестник Академии военных наук*. 2023. № 1 (82). С. 31–35.
13. Противодействие фальсификации истории великой отечественной войны / Бочарников И.В., Суздалева Т.Р., Федоров К.В., Криворучко А.А., Петренко А.И., Зеленков М.Ю., Кандыбович С.Л., Разина Т.В., Овсянникова О.А., Трипольский В.Б. Москва, 2020.