

Данная статья представляет собой расширенную версию доклада, представленного на [конференции GraphiCon 2025](#).

Информационная плотность объектов в цифровой среде: теоретические основы

Н.А. Бондарева¹, А.Е. Бондарев², С.В. Андреев³, И.Г. Рыжова⁴

Институт прикладной математики им. М.В. Келдыша Российской академии наук

¹ ORCID: 0000-0002-7586-903X, nicibond9991@gmail.com

² ORCID: 0000-0003-3681-5212, bond@keldysh.ru

³ ORCID: 0000-0001-8029-1124, esa@keldysh.ru

⁴ ORCID: 0000-0003-1613-3038, ryzhova@gin.keldysh.ru

Аннотация

В статье представлена теоретическая концепция оценки информационной плотности объектов в цифровой среде. Проведен анализ ограничений существующих методов количественной оценки информационного пространства, основанных преимущественно на простых метриках объема данных. Предложена концепция информационного поля объекта как совокупности всех информационных единиц, содержащих упоминания исследуемого объекта в цифровом пространстве.

Рассмотрены перспективы применения методологии в информационной безопасности, анализе социальных медиа и подготовке данных для обучения нейронных сетей. Предложенный подход открывает новые возможности для комплексной оценки информационных ресурсов и может найти применение в поисковых системах, рекомендательных алгоритмах и системах анализа больших данных.

Ключевые слова: информационная плотность, цифровая среда, информационное пространство, анализ данных, информационная аналитика.

1. Введение

Современное цифровое пространство характеризуется экспоненциальным ростом объемов информации и качественным усложнением структуры информационных потоков. Согласно различным оценкам, ежедневно в мировом информационном пространстве генерируется свыше двух с половиной квинтиллионов байт данных [1], значительная часть которых представляет собой неструктурированную информацию в виде текстового контента, мультимедийных материалов и метайнформационных структур. В условиях такого информационного изобилия перед исследователями и практиками встает фундаментальный вопрос о разработке адекватных методов оценки информационной полноты и качественных характеристик данных, доступных относительно интересующих объектов реального мира.

Традиционные подходы к количественной оценке информационных ресурсов, опирающиеся на простые метрики объема данных, частоты упоминаний или статистических показателей встречаемости, демонстрируют существенную ограниченность при решении задач комплексной характеристики информационной ценности агрегированных данных. Существующий арсенал методов библиометрического и наукометрического анализа [2], несмотря на развитый математический аппарат для оценки научного влияния публикаций и цитирования, не предоставляет инструментария для комплексной оценки информационного окружения произвольных объектов, выходящих за рам-

ки научно-академической сферы. Аналогично, методологический арсенал анализа социальных сетей, ориентированный на оценку популярности и влияния узлов в сетевых структурах, не учитывает качественные параметры информационного содержания и не решает задачу оценки уникальности информационных единиц.

Особую остроту данная проблематика приобретает в контексте интенсивного развития систем искусственного интеллекта и технологий машинного обучения. Эффективность функционирования современных нейросетевых архитектур детерминирована качественными характеристиками обучающих данных, включающими не только их количественную полноту, но и информационную достоверность, репрезентативность и отсутствие избыточности. Однако доминирующие в настоящее время подходы к формированию обучающих датасетов концентрируются преимущественно на количественных аспектах выборочных совокупностей, таких как объем выборки и сбалансированность представленности различных классов объектов, оставляя без должного внимания качественную оценку информационного насыщения данных относительно исследуемых сущностей.

Аналогичные методологические задачи возникают в смежных областях информационных технологий, включающих системы информационного поиска и извлечения знаний, рекомендательные системы, технологии автоматической обработки естественного языка и построения семантических графов знаний. Указанные системы, функциональное предназначение которых состоит в агрегации, структурировании и интеллектуальном анализе информации о реальных объектах различной природы, включающих организационные структуры, события и абстрактные концепции, испытывают потребность в разработке инструментов для адекватной оценки полноты и качественных характеристик доступной информации.

Исследовательские работы в предметной области разрешения сущностей и построения графов знаний развивают методологические подходы к связыванию и агрегации разнородной информации об объектах исследования [3-4], однако концентрируют внимание преимущественно на технических аспектах интеграции гетерогенных данных, не предлагая решений для задачи оценки информационной плотности и качественных характеристик получаемых объектных представлений.

Научные исследования в области информационного поиска и извлечения информации выработали развитый математический аппарат для оценки релевантности документов пользовательским запросам и информационным потребностям, включающий классические метрики точности и полноты поиска, однако существующие методы не решают концептуально отличную задачу комплексной оценки информационного окружения объекта как целостного феномена цифрового пространства.

В данной работе предлагается концептуальная модель анализа информационного пространства, основанная на введении понятия плотности информационного поля объекта. Информационное поле объекта определяется как совокупность всех информационных единиц, содержащих упоминания, описания или ссылки на исследуемый объект и доступных в открытых цифровых источниках. Принципиальной характеристикой такого информационного поля является его плотность, представляющая собой интегральную меру, отражающую не только количественные параметры доступной информации, но и ее существенные качественные характеристики, включающие уникальность информационного содержания, релевантность относительно объекта исследования, авторитетность информационных источников и временную актуальность представленных данных.

Основной целью данного исследования является разработка теоретических основ концепции плотности информационного поля и формулирование методологического аппарата для ее практического применения в задачах анализа и комплексной оценки информационных ресурсов цифрового пространства. Достижение поставленной цели предполагает решение комплекса взаимосвязанных исследовательских задач, включающих формулирование строгих определений базовых понятий информационного поля

объекта, плотности информационного поля и информационного кванта, разработку математической модели для расчета плотности информационного поля с учетом многомерных характеристик информационного содержания.

2. Анализ существующих подходов к оценке информационных ресурсов

Современная научная литература содержит обширный набор исследований, посвященных различным аспектам оценки и анализа информационных ресурсов, однако существующие подходы демонстрируют фрагментарность и ограниченную применимость к задаче комплексной оценки информационного окружения произвольных объектов реального мира.

Фундаментальные теоретические основы количественной оценки информации были заложены в классических работах Клода Шеннона по математической теории информации, где информационное содержание сообщения определяется через энтропийные характеристики источника и канала передачи. Шенноновская концепция информационной энтропии [5] предоставляет собой математический аппарат для измерения неопределенности и информационной емкости систем, однако данный подход оперирует исключительно синтаксическими характеристиками информации, не учитывая семантическое содержание сообщений и их прагматическую ценность для решения конкретных задач. Развитие данного направления в работах Андрея Колмогорова по алгоритмической теории информации [6] вводит понятие сложности строки как длины кратчайшей программы, способной ее воспроизвести, что позволяет формализовать интуитивные представления о содержательности и неслучайности информационных последовательностей.

Существенный массив современных исследований сосредоточен на проблематике оценки качества данных [7], рассматриваемой как многомерная характеристика информационных ресурсов. Концептуальные модели качества данных, развиваемые в работах исследователей данного направления, включают такие фундаментальные измерения, как точность информационного содержания, отражающая соответствие данных реальному состоянию описываемых объектов, полнота представления, характеризующая степень покрытия релевантных аспектов предметной области, согласованность данных, определяющая отсутствие внутренних противоречий в информационных структурах, актуальность сведений относительно временного контекста их использования, и релевантность информации применительно к поставленным аналитическим задачам. Указанные исследования предлагают разнообразные метрики и алгоритмические подходы к автоматической оценке качественных характеристик данных, включая статистические методы выявления аномалий и несогласованностей, техники сравнительного анализа множественных источников для верификации фактической точности, и эвристические алгоритмы оценки полноты на основе структурных особенностей предметных областей. Вместе с тем, существующие исследования качества данных демонстрируют методологическую ориентацию преимущественно на структурированные информационные ресурсы, такие как реляционные базы данных и формализованные каталоги, не предлагая адекватных решений для оценки качественных характеристик неструктурированной информации, естественным образом агрегирующейся вокруг объектов в открытом цифровом пространстве.

Значительное исследовательское внимание в последние десятилетия привлекают методы разрешения сущностей и построения семантических графов знаний, направленные на интеграцию и структурирование разнородной информации об объектах реального мира. Задача разрешения сущностей, состоящая в идентификации и связывании различных упоминаний одного объекта в гетерогенных информационных источниках, решается посредством комбинации методов сравнения строковых представлений, анализа контекстуального окружения упоминаний, и машинного обучения на

размеченных корпусах данных. Современные подходы к построению графов знаний, такие как реализованные в проектах Google Knowledge Graph, DBpedia и Wikidata [8-10], демонстрируют впечатляющие результаты в агрегации структурированной информации о миллионах сущностей различных типов, предоставляя унифицированные интерфейсы для доступа к интегрированным знаниям. Однако существующие исследования в данной области концентрируются преимущественно на технических аспектах интеграции и верификации структурированных данных, не предлагая методологических решений для оценки информационной плотности и качественных характеристик неструктурированного информационного окружения объектов, которое не может быть адекватно представлено в формате структурированных графовых отношений.

Классические исследования в области информационного поиска и извлечения информации выработали развитый математический аппарат для оценки эффективности поисковых систем и релевантности документов пользовательским запросам. Фундаментальные метрики точности и полноты поиска, а также их гармоническое среднее в виде F-меры, предоставляют количественные инструменты для оценки качества функционирования информационно-поисковых систем в контролируемых экспериментальных условиях. Развитие данного направления привело к созданию сложных моделей релевантности, учитывающих множественные факторы, включающие текстовое сходство между запросом и документом, авторитетность источников информации, временные характеристики документов, пользовательское поведение и контекстуальные особенности информационных потребностей. Современные подходы к ранжированию документов в веб-поиске, основанные на алгоритмах типа PageRank [11] и его модификациях, демонстрируют эффективность учета структурных характеристик информационного пространства, таких как паттерны гиперссылок и цитирования, для оценки авторитетности и значимости информационных ресурсов. Тем не менее, существующие методы информационного поиска решают концептуально иную задачу оценки соответствия отдельных документов конкретным информационным запросам, не предлагая подходов к комплексной оценке информационного окружения объектов как целостных феноменов цифрового пространства, характеризующихся специфическими паттернами информационной агрегации и качественного разнообразия источников.

Библиометрические и наукометрические исследования развили изоцированные методы оценки научного влияния и значимости публикаций, основанные на анализе паттернов цитирования и кооперативных связей между исследователями. Классические индикаторы, такие как индекс Хирша и его многочисленные модификации, предоставляют комплексные оценки продуктивности и влияния ученых, учитывающие как количественные характеристики публикационной активности, так и качественные параметры, отражаемые в частоте цитирования работ научным сообществом. Современные подходы к наукометрическому анализу включают в себя сетевые методы исследования научных коллабораций, временной анализ эволюции исследовательских направлений, и междисциплинарные метрики, отражающие широту влияния научных результатов. Несмотря на методологическую развитость библиометрических подходов, их применимость ограничивается специфическим контекстом научно-академической деятельности и не распространяется на оценку информационных характеристик произвольных объектов, не связанных с формальными системами научного цитирования и экспертной оценки.

Исследования в области семантического веба и онтологического инжиниринга предлагают формализованные подходы к представлению и структурированию знаний о предметных областях, основанные на логических формализмах и стандартизированных языках описания ресурсов. Онтологические модели предоставляют выразительные средства для спецификации концептуальных структур предметных областей, включая иерархии классов объектов, свойства и отношения между сущностями, ограничения целостности и правила логического вывода [12-13]. Технологии связанных данных, развиваемые в рамках инициативы семантического веба, демонстрируют практиче-

скую эффективность интеграции распределенных информационных ресурсов посредством стандартизированных протоколов и форматов представления данных. Методы оценки качества онтологий включают анализ логической согласованности концептуальных моделей, оценку полноты покрытия предметных областей, и метрики соответствия онтологических структур реальным характеристикам описываемых феноменов. Вместе с тем, онтологические подходы предполагают наличие экспертно разработанных концептуальных моделей и не решают задачу оценки качественных характеристик спонтанно формирующейся неструктурированной информации, которая не может быть адекватно описана в рамках предзаданных онтологических схем.

Анализ существующих исследовательских направлений выявляет принципиальный пробел в методологическом арсенале современных информационных технологий, связанный с отсутствием концептуальных подходов к комплексной оценке информационных полей, естественным образом формирующихся вокруг произвольных объектов в неструктурированном цифровом пространстве. Существующие методы, несмотря на их развитость в рамках специализированных предметных областей, не предоставляют адекватных инструментов для интегральной оценки качественных характеристик разнородной информации, агрегирующейся вокруг объектов различной природы в условиях информационного изобилия современного цифрового пространства. Данный методологический вакуум обуславливает необходимость разработки новых концептуальных подходов, способных обеспечить теоретические основы для комплексной оценки информационной плотности и качественных характеристик объектно-ориентированных информационных полей.

3. Теоретические основы концепции плотности информационного поля

Разработка концептуального аппарата для анализа информационных полей объектов требует формулирования новых теоретических конструктов, которые позволили бы преодолеть ограничения существующих подходов к оценке информационных ресурсов. Предлагаемая теоретическая модель основывается на синтезе идей классической теории информации, современных представлений о семантическом пространстве и принципов системного анализа сложных информационных структур. Центральной предпосылкой данного подхода является понимание того, что информация об объектах реального мира не существует в цифровом пространстве в виде изолированных дискретных единиц, но формирует сложные взаимосвязанные структуры, которые могут быть концептуализированы как информационные поля, обладающие специфическими топологическими и качественными характеристиками.

Концепция информационного поля объекта основывается на метафорическом переносе физических представлений о полевых структурах в область анализа информационных явлений. Подобно тому, как физические поля характеризуются распределением энергии или вещества в пространстве, информационное поле объекта представляет собой распределение информационных единиц в цифровом пространстве, где каждая информационная единица обладает определенной "массой" или значимостью, а совокупность таких единиц формирует сложную топологическую структуру с различными зонами концентрации и разреженности информации. Данная концептуализация позволяет применить к анализу информационных структур математический аппарат, разработанный для изучения полевых явлений, включая понятия плотности, градиента, потока и других характеристик, описывающих пространственное распределение физических величин.

Формальное определение информационного поля объекта может быть сформулировано следующим образом: информационное поле объекта O представляет собой множество всех информационных единиц $I = \{i_1, i_2, \dots, i_n\}$, доступных в цифровом пространстве и содержащих прямые или косвенные упоминания, описания, ссылки или

любые иные формы информационной связи с данным объектом. Математически это может быть выражено как $IF(O) = \{i \in I \mid R(i, O) > \theta\}$, где $R(i, O)$ представляет функцию релевантности информационной единицы i относительно объекта O , а θ является пороговым значением, определяющим минимальную степень релевантности для включения информационной единицы в поле объекта. Важно отметить, что границы информационного поля носят размытый характер, поскольку релевантность информации может варьироваться в широких пределах, от прямых упоминаний объекта до сложных контекстуальных ассоциаций, установление которых требует глубокого семантического анализа.

Информационное поле объекта характеризуется рядом фундаментальных свойств, определяющих его структуру и динамические характеристики. Пространственная неоднородность поля проявляется в том, что различные области цифрового пространства содержат неравномерное распределение информации об объекте, при этом некоторые источники и платформы могут аккумулировать значительные объемы релевантной информации, в то время как другие сегменты информационного пространства практически не содержат упоминаний о данном объекте. Временная динамичность представляет еще одно ключевое свойство информационных полей, поскольку они непрерывно эволюционируют под воздействием появления новых информационных источников, обновления существующих материалов, изменения популярности объекта и других факторов, влияющих на информационную активность вокруг него. Многомерность информационного поля обусловлена тем, что информация об объекте может существовать в различных форматах и на различных семантических уровнях, включая фактографические данные, аналитические материалы, мультимедийный контент, метаинформацию и контекстуальные связи.

Центральной характеристикой информационного поля является его плотность, которая представляет собой интегральную меру концентрации качественной информации об объекте в единице информационного пространства. В отличие от простых количественных метрик, таких как общий объем данных или частота упоминаний, плотность информационного поля должна учитывать качественные параметры информационного содержания, включающие уникальность информации, ее релевантность относительно объекта исследования, авторитетность источников и временную актуальность данных. Концептуально плотность информационного поля может быть понимаема как мера информационной "массы" объекта в цифровом пространстве, отражающая не только количество доступной информации, но и ее способность формировать полное и точное представление об объекте исследования. В более общем смысле плотность информационного поля характеризует способность технологии, такой как искусственный интеллект, воссоздавать образ объекта на основе собранных данных [14], установленных паттернов и взаимосвязей.

Стоит отметить, что данная концепция представляет собой попытку отойти от простого количественного измерения информации в битах или количестве тегов, ссылок или упоминаний, так как в условиях существования современных языковых моделей особую ценность информация приобретает, когда при её интерпретации и поиске можно оперировать не просто словами, а смысловыми конструкциями и контекстами. С точки зрения количественного измерения информации, бытует традиционная установка: чем больше данных доступно, тем точнее и полнее можно воссоздать цифровой образ объекта. Однако разнородность информации, касающейся одного и того же объекта исследования в контексте разных трактовок, делает идею создать для всех нейтральную модель, ответ которой устроит каждую сторону, практически невозможной. Как говорится в отчёте NewsGuard [15-16], рост информационного охвата и подключение веб-поиска к языковым моделям и чат-ботам обернулись заметным ухудшением точности выдаваемых результатов, а во время событий, происходящих «онлайн», чаты чаще усиливали ложные нарративы, подтягивая материалы из сомнительных ис-

точников и не делая разницы между авторитетными изданиями и их пропагандистскими двойниками.

Вместе с тем, интегральная оценка плотности информации как истинной, так и заведомо ложной, окружающей объект, может позволить проанализировать объем источников, на которые потенциально может опираться модель, а также какой образ объекта она с большей вероятностью сформирует, какие паттерны будет использовать и какое мнение изберет наиболее приоритетным. На фоне того, что в свете развития языковых моделей в разных политических полюсах рынок, вероятно, будет подталкивать ИИ-сервисы к более ярко выраженной позиции в попытке удовлетворить ожидания своей аудитории, возникает необходимость в создании нейтральных и объективных критериев оценки объемов информации, чтобы иметь возможность интерпретировать основные направления развития цифровых образов тех или иных явлений, на которые будут ориентироваться языковые модели. В недалеком будущем развитие таких сервисов может привести к тому, что разные языковые модели, ограниченные по кардинально различающимся принципам, будут давать диаметрально противоположные ответы на один и тот же вопрос. В свете этого собрать относительно объективную информационную картину с каждым годом будет всё труднее.

Плотность информационного поля объекта может характеризоваться количеством условных единиц (условная единица — это квант уникальной, неповторяющейся дословно информации, содержащей в себе мысль/анализ/исследование/факт/вывод, касающийся искомого объекта). Тогда условно 1000 сгенерированных мыслей, касающихся объекта, повысят плотность его информационного поля, а, к примеру, 1000 репостов (1000 упоминаний одного и того же термина без малейшего изменения формулировки) не изменят уровень плотности. В этом случае количество информации в традиционном понимании информатики, т.е. измеряемое в битах может быть только одним из параметров, но не самым значимым.

В связи с этим можно ввести название для такой условной единицы. Инфон (infon, от "information" + суффикс "on") — в первичном определении это неповторяющиеся единицы информации, содержащие оригинальные мысли, анализ, исследования или факты об объекте.

Однако каждая условная единица (квант информации) так или иначе, должна иметь свой "вес" или "значимость". К примеру, научная статья с оригинальным исследованием объекта и случайный комментарий в социальной сети, хотя оба содержат уникальную информацию, имеют разную ценность для формирования информационного поля. Таким образом, более "тяжелые" информационные единицы имеют более сильное влияние в информационном поле. Также важно учитывать временную составляющую. Информация имеет свойство устаревать, и её значимость может меняться со временем. Например, научная статья десятилетней давности может иметь меньший вес, чем свежее исследование, если только она не является фундаментальной для данной области.

Интересным дополнением может стать концепция "информационных резонансов". Когда несколько независимых источников подтверждают одну и ту же информацию разными способами (не простое копирование), это может создавать эффект усиления — подобно тому, как волны могут усиливать друг друга при резонансе. Такие резонансы могут существенно и резко повышать плотность информационного поля в определенные моменты времени. Визуализация подобных временных колебаний может существенно улучшить понимание развития того или иного явления.

Важным аспектом является также связность таких информационных квантов. Отдельные фрагменты информации, связанные между собой логическими или причинно-следственными связями, могут формировать более устойчивые и значимые структуры в информационном поле, чем разрозненные факты.

Если рассматривать инфон с точки зрения этих аспектов, то его определение уточняется: инфон — это минимальная неделимая единица уникальной информации об объекте, которая не может быть редуцирована без потери смыслового содержания. Инфон

представляет собой минимальную уникальную единицу информации, содержащую законченную мысль факт об объекте, которая не является прямым повторением существующей информации. Здесь возможно двоякое толкование в модели: можно рассматривать инфоны, как единицу измерения, и тогда они должны быть одинаковыми. Из этого следует необходимость обусловить, что же является минимальной единицей уникальной информации.

Другой подход к модели допускает, что каждый инфон определяется как квант уникальной информации и может иметь свой информационный вес и релевантность (рис. 1). Инфон характеризуется атомарностью, означающей невозможность его дальнейшего деления без потери семантической и информационной целостности, уникальностью, предполагающей отсутствие точных дубликатов данной информационной единицы в других источниках, релеванностью, обеспечивающей прямую связь содержания инфона с объектом исследования, и верифицируемостью, позволяющей осуществить проверку фактической достоверности содержащейся в инфоне информации.

На рисунке 1 приведен пример возможной визуализации такого явления, когда представленный объект окружен разнородными частицами, имеющими градацию по цвету и размеру. Размер каждой частицы отражает размер кванта информации, цвет демонстрирует меру концептуальной связи с объектом от непосредственно относящейся к нему, до имеющей весьма отдаленное отношение.

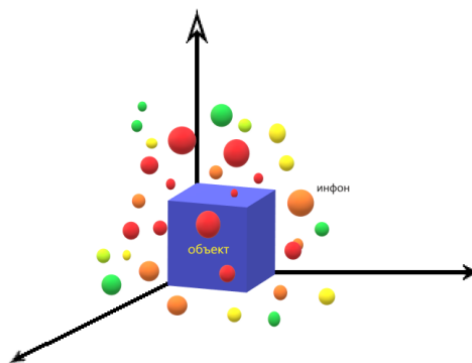


Рис. 1. Пример визуализации плотности информационного поля объекта

Процесс выделения инфонов из общего массива информационного поля представляет собой сложную аналитическую задачу, где требуется комбинирование автоматических методов обработки естественного языка и экспертного анализа содержания информационных материалов. Алгоритмическая идентификация может основываться на методах сегментации текстового содержания с выделением семантически завершенных фрагментов, анализе новизны информации посредством сравнения с существующими базами знаний, оценке фактической уникальности через детекцию плагиата и дублирования, и определении релевантности с использованием методов семантического анализа и машинного обучения. Каждый идентифицированный инфон может быть охарактеризован набором количественных параметров, включающих меру его уникальности, степень релевантности объекту, оценку достоверности содержащейся информации, и показатель информационной ценности для формирования целостного представления об объекте.

Поскольку информация в современном мире представлена в самых разных видах (визуальная, аудио, видеопоток, текст, комментарии в соцсетях, наборы численных данных), отдельной задачей становится объединение их в единую структуру. Ключ к объединению разных типов информации лежит в концепции многомерного информационного пространства. Информационное поле можно представить как многослойную структуру, где каждый тип данных формирует свой слой, но все они взаимосвязаны и влияют друг на друга. Это похоже на то, как в нейронных сетях различные модальности

(текст, изображения, звук) могут быть преобразованы в единое векторное пространство.

Для унификации различных типов данных подходит концепция "информационных эмбедингов" – преобразование любого типа информации в универсальное векторное представление [17]. Современные технологии уже позволяют это делать: CLIP может находить связи между текстом и изображениями, wav2vec преобразует звук в векторы, а большие языковые модели превращают текст в многомерные представления. В этом случае важно учитывать "коэффициент информационной плотности" для разных типов данных. Например, одна секунда видео может нести больше информации об объекте, чем текстовое описание той же продолжительности, но текстовая аналитическая статья может содержать более глубокую семантическую информацию, чем простая фотография.

Также возникает вопрос "перекрестной валидации" между разными типами данных. Если информация из разных источников и форматов подтверждает друг друга, это повышает достоверность каждого отдельного информационного кванта. Например, если текстовое описание события подкрепляется видеозаписью и числовыми данными, общая достоверность информации возрастает. В контексте искусственного интеллекта такой подход открывает новые возможности для создания мультимодальных систем, способных формировать целостное представление об объектах на основе разнородных данных. Это может оказаться ещё одной системой организации данных, полезной для развития систем общего искусственного интеллекта, которые должны уметь работать с информацией целостно, как это делает человеческий мозг.

В настоящее время большие генеративные модели искусственного интеллекта склонны накапливать множество отдельных алгоритмических правил — специализированных схем для конкретных случаев, которые не объединяются в целостную систему знаний. Такие локальные закономерности часто противоречат друг другу, создавая внутренние конфликты в работе системы. Исследовательские попытки найти в структуре моделей цельные концептуальные представления [18] выявляют только разрозненные информационные фрагменты, не образующие единого связного образа. Тем не менее, такие распределенные правила имеют определенную практическую ценность. Огромные параметрические возможности языковых моделей позволяют хранить подобные паттерны в больших количествах, и часто количество компенсирует отсутствие четкой структуры. Возможность создания проверяемых внутренних представлений открывает пути для борьбы с искусственными галлюцинациями, повышения надежности логических выводов и обеспечения большей прозрачности работы интеллектуальных систем.

Математическая формализация плотности информационного поля может быть представлена в виде взвешенной суммы информационных вкладов отдельных элементов поля, где каждый элемент (инфон) оценивается по множественным критериям качества. Пусть информационное поле (IF – Information Field) объекта представляет собой многомерное векторное пространство. Базовая формула могла бы выглядеть так:

$$IF = \sum (Q_i \times W_i \times T_i \times R_i) \times K_i$$

где Q_i – некий информационный квант, W_i является весовым коэффициентом, отражающим тип и значимость информационной единицы i , $T(i)$ учитывает временную актуальность данных, $R(i)$ определяет релевантность относительно объекта, K является неким нормализующим коэффициентом.

Также формула может быть дополнена такими параметрами, как $A(i)$, который отражает авторитетность источника информации, и $U(i)$, характеризующий уникальность информационного содержания.

Каждый из этих коэффициентов имеет свою методику расчета, основанную на объективных параметрах и существующих методах анализа данных. Некоторые из этих коэффициентов имеет смысл искать и заимствовать из существующих исследований по

big data. Существует ряд смежных областей исследований: информационная теория поля в физике [19], семантические пространства в лингвистике [20 ссылка на Клышинского/упругие карты], цифровая экология информации и квантовая теория информации. Эти направления предоставляют полезные инструменты и методологии, которые можно адаптировать для развития теории информационных полей.

К примеру, если рассмотреть W_i (весовой коэффициент источника), то ближайшими аналогами можно определить такие сущности, как импакт-фактор научных журналов, PageRank от Google и индекс цитирования. Возможная формула расчета может быть примерно такой:

$$W_i = (A_s \times C_s \times R_s \times V_s) / N_{max}$$

где A_s (Authority Score) является авторитетностью источника (0-1), C_s (Citation Score) отвечает за индекс цитируемости, R_s (Reliability Score) характеризует показатель надежности на основе исторических данных, V_s (Verification Score) – возможность верификации информации, а N_{max} – некий нормализующий максимум

Также для показателя T_i можно обратиться к реально существующим примерам: научные базы данных учитывают "возраст" публикаций при расчете их значимости, а Netflix использует похожие временные коэффициенты для ранжирования контента. Метрики для оценки релевантности информации R_i используются в современных поисковых системах и системах обработки естественного языка (NLP).

Нормализующий коэффициент определяется эмпирически для конкретной предметной области и может рассчитываться как:

$$K = 1 / \max(IF)$$

где $\max(IF)$ - максимально возможное значение информационного поля в данной области.

Взаимодействие между концептами информационного поля, его плотности и составляющих инфонов формирует теоретическую модель, позволяющую осуществлять качественный анализ информационных структур в цифровом пространстве. Инфоны выступают в роли отдельных частиц информационного поля, их совокупность определяет структурные характеристики поля, а их качественные параметры детерминируют общую плотность информационного поля объекта. Данная модель предоставляет теоретические основы для разработки практических методов оценки информационных ресурсов, позволяя перейти от интуитивных представлений о "богатстве" или "бедности" информации об объекте к более строгим количественным оценкам, основанным на анализе качественных характеристик информационного содержания.

4. Перспективы применения и развития теории

Предложенный подход к оценке информационной плотности в цифровой среде открывает новые возможности для практического применения и дальнейшего развития. В числе основных направлений потенциального использования методологии и перспективы её развития можно назвать такие области, как: информационная безопасность, анализ социальных медиа и контента, подготовка данных для обучения нейронных сетей.

В сфере информационной безопасности методология оценки информационной плотности может найти применение в нескольких ключевых направлениях. Во-первых, анализ плотности информационного поля может позволить выявлять аномальные всплески активности, которые могут свидетельствовать о целенаправленных информационных кампаниях или атаках. Во-вторых, оценка качественных характеристик информационного пространства в перспективе поможет идентифицировать источники недостоверной информации и отслеживать распространение дезинформации. Также может оказаться актуальной задача раннего обнаружения информационных угроз путем анализа динамики изменения информационной плотности.

В контексте анализа социальных медиа предложенная методология может предоставить инструментарий для более глубокого понимания информационных процессов. Оценка информационной плотности позволяет выявлять значимые тренды и отделять их от информационного шума, что особенно важно в условиях перенасыщенности социальных сетей контентом. Анализ качественных характеристик информационного поля помогает определять реальное влияние контента и его авторов, выходя за рамки простых количественных метрик, таких как число лайков или репостов.

В свете развития данной теории можно выявить следующие направления дальнейших исследований.

1. Развитие математического аппарата представляет собой одно из ключевых направлений. Необходима разработка более точных моделей для описания взаимодействия различных компонентов информационного поля, создание методов количественной оценки качественных характеристик информации. Особое внимание следует уделить созданию математических моделей, учитывающих временную динамику информационных процессов и нелинейный характер взаимодействия различных типов информации.

2. Экспериментальная проверка методологии требует проведения серии исследований в различных предметных областях. Необходимо подтверждение эффективности предложенных методов оценки информационной плотности на реальных данных, валидация результатов в различных контекстах применения. Важным аспектом является разработка стандартизированных методик проведения экспериментов и критериев оценки результатов.

3. Создание практических инструментов является необходимым этапом для широкого внедрения методологии. Требуется разработка программного обеспечения, способного автоматизировать процессы анализа информационной плотности, создание удобных интерфейсов для работы с данными, интеграция с существующими системами анализа информации.

Развитие предложенной методологии может оказать существенное влияние на несколько аспектов информационных технологий. В области поисковых систем появляется возможность создания более точных алгоритмов ранжирования результатов, учитывающих не только количественные, но и качественные характеристики информации. В сфере искусственного интеллекта методология может способствовать развитию более совершенных систем обработки естественного языка и анализа данных. Также подобная методология может внести вклад в развитие персонализированных рекомендательных систем, способных более точно учитывать контекст и качество информации.

5. Выводы и заключение

Представленное исследование носит концептуальный характер и направлено на формулирование теоретических основ нового подхода к анализу информационного пространства через призму концепции плотности информационного поля. Основным результатом работы является введение системы взаимосвязанных понятий, включающих информационное поле объекта, его плотность и составляющие информационные кванты (инфоны), которые в совокупности образуют целостную концептуальную модель для качественной оценки информационных ресурсов в цифровом пространстве.

Предложенные в рамках исследования математические формализации имеют преимущественно иллюстративный характер и служат для демонстрации принципиальной возможности количественного описания качественных характеристик информационных структур. Дальнейшая разработка строгого математического аппарата потребует проведения обширных эмпирических исследований для определения конкретных параметров функций уникальности, релевантности, авторитетности и других компонентов предложенной модели плотности. Особого внимания заслуживает задача операционализации концепта инфона, требующая разработки алгоритмических процедур для

автоматического выделения информационных квантов из неструктурированных текстовых массивов и их качественной оценки.

Теоретическая значимость предлагаемого подхода заключается в преодолении ограничений существующих методов оценки информационных ресурсов, которые традиционно фокусируются либо на количественных характеристиках данных, либо на узкоспециализированных аспектах информационного качества. Практическое значение разработанной концепции определяется ее потенциальными применениями в развитии систем искусственного интеллекта нового поколения, способных формировать более надежные и верифицируемые внутренние репрезентации объектов реального мира. Переход от статистических закономерностей, извлекаемых из неконтролируемых текстовых корпусов, к систематическому анализу информационной плотности может способствовать существенному снижению частоты генерации недостоверной информации в языковых моделях и повышению прозрачности процессов принятия решений в интеллектуальных системах. Представленная концептуальная модель, таким образом, может служить отправной точкой для формирования новой исследовательской парадигмы в области анализа информационных ресурсов и их качественных характеристик.

Список литературы

1. Austin Harris, 2.5 quintillion bytes of data are produced by people every day, 2021, - URL: <https://appdeveloperomagazine.com/2.5-quintillion-bytes-of-data-are-produced-by-people-every-day/>
2. Малахов В.А., Библиометрический анализ как метод науковедческих исследований: возможности и ограничения // Науковедческие исследования. 2022. №1. URL: <https://cyberleninka.ru/article/n/bibliometricheskiy-analiz-kak-metod-naukovedcheskih-issledovaniy-vozmozhnosti-i-ogranicheniya> (дата обращения: 08.09.2025).
3. Hogan A. et al. Knowledge graphs //ACM Computing Surveys (Csur). – 2021. – Т. 54. – №. 4. – С. 1-37.
4. Кислицына М.Ю. Analysis of the Error Structure in Identifying the Author of a Text Using the Nearest Neighbor Graphs (2025). Научная визуализация 17.2: 110 - 122, DOI: 10.26583/sv.17.2.08
5. Shannon C.E. A Mathematical Theory of Communication (1948)
6. Kolmogorov, A.N. (1965). "Three approaches to the quantitative definition of information"
7. Batini C. et al. Data and information quality //Cham, Switzerland: Springer International Publishing. – 2016. – Т. 63.
8. Singhal A. Introducing the Knowledge Graph: things, not strings (2012) URL: <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
9. Dbpedia: Global and Unified Access to Knowledge Graphs, URL: <https://www.dbpedia.org/>
10. Wikidata, URL: https://www.wikidata.org/wiki/Wikidata:Main_Page
11. Brin S., Page L. The anatomy of a large-scale hypertextual web search engine //Computer networks and ISDN systems. – 1998. – Т. 30. – №. 1-7. – С. 107-117.
12. S.I. Chuprina, I.A. Labutin. A High-Level Adaptation Toolkit for Unified Integration of Software Systems with Neural Interfaces (2024). Научная визуализация 16.4: 11 - 24, DOI: 10.26583/sv.16.4.02
13. S.I. Chuprina. Using Data Fabric Architecture to Create Personalized Visual Analytics Systems in the Field of Digital Medicine (2023). Научная визуализация 15.5: 50 - 63, DOI: 10.26583/sv.15.5.05
14. N.A. Bondareva. The Impact of Input Data Density on the Performance of Graphic Neural Networks (2024). Scientific Visualization 16.5: 109 - 119, DOI: 10.26583/sv.16.5.08

15. Morrone M., Exclusive: Popular chatbots amplify misinformation, 2025, - URL: https://www.axios.com/2025/09/04/popular-chatbots-amplify-misinformation?utm_source=Securitylab.ru
16. AI False Information Rate Nearly Doubles in One Year, 2025, - URL: <https://www.newsguardtech.com/ai-monitor/august-2025-ai-false-claim-monitor/>
17. Li S., Guo H., Tang X., Tang R., Hou L., Li R., Zhang R. Embedding Compression in Recommender Systems: A Survey (2024), arXiv:2408.02304
18. Vafa K., Chen J. Y., Rambachan A., Kleinberg J., Mullainathan S. Evaluating the World Model Implicit in a Generative Model (2024) <https://doi.org/10.48550/arXiv.2406.03689>
19. Torsten A. Enßlin. Information theory for fields (2018). Annalen der Physik 2019, vol. 531, issue 3, p. 1800127 DOI: 10.1002/andp.201800127
20. Бондарев А.Е., Бондаренко А.В., Галактионов В.А., Клышинский Э.К. Визуальный анализ кластерных структур в многомерных объемах текстовой информации / Научная визуализация, Т.8, № 3, 2016, с. 1-24

Information density of objects in digital environment: theoretical foundations

N.A. Bondareva¹, A.E. Bondarev², S.V. Andreev³, I.G. Ryzhova⁴

Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

¹ ORCID: 0000-0002-7586-903X, nicibond9991@gmail.com

² ORCID: 0000-0003-3681-5212, bond@keldysh.ru

³ ORCID: 0000-0001-8029-1124, esa@keldysh.ru

⁴ ORCID: 0000-0003-1613-3038, ryzhova@gin.keldysh.ru

Abstract

The paper presents a theoretical concept for evaluating information density of objects in digital environment. An analysis of limitations of existing methods for quantitative assessment of information space, based predominantly on simple data volume metrics, has been conducted. The concept of an object's information field is proposed as an aggregate of all informational units containing mentions of the studied object in digital space.

The prospects for applying this methodology in information security, social media analysis, and data preparation for neural network training are examined. The proposed approach opens new opportunities for comprehensive evaluation of information resources and may find application in search engines, recommendation algorithms, and big data analysis systems.

Keywords: information density, digital environment, information space, data analysis, information analytics.

References

1. Austin Harris, 2.5 quintillion bytes of data are produced by people every day, 2021, - URL: <https://appdeveloperomagazine.com/2.5-quintillion-bytes-of-data-are-produced-by-people-every-day/>
2. Malakhov V.A., Bibliometric analysis as a method of science studies: possibilities and limitations // Science studies. 2022. No. 1. URL: <https://cyberleninka.ru/article/n/bibliometricheskiy-analiz-kak-metod-naukovedcheskih-issledovaniy-vozmozhnosti-i-ogranicheniya> (date of access: 09/08/2025).
3. Hogan A. et al. Knowledge graphs //ACM Computing Surveys (Csur). – 2021. – T. 54. – No. 4. – S. 1-37.
4. Kislitsyna M. Yu . Analysis of the Error Structure in Identifying the Author of a Text Using the Nearest Neighbor Graphs (2025). Scientific Visualization 17.2: 110–122, DOI: 10.26583/sv.17.2.08
5. Shannon C.E. A Mathematical Theory of Communication (1948)
6. Kolmogorov, A. N. (1965). "Three approaches to the quantitative definition of information"
7. Batini C. et al. Data and information quality //Cham, Switzerland: Springer International Publishing. – 2016. – T. 63.
8. Singhal A. Introducing the Knowledge Graph: things, not strings (2012) URL: <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
9. Dbpedia : Global and Unified Access to Knowledge Graphs, URL: <https://www.dbpedia.org/>
10. Wikidata, URL: https://www.wikidata.org/wiki/Wikidata:Main_Page

11. Brin S., Page L. The anatomy of a large-scale hypertextual web search engine //Computer networks and ISDN systems. – 1998. – T. 30. – No. 1-7. - WITH . 107-117.
12. SI Chuprina, IA Labutin . A High-Level Adaptation Toolkit for Unified Integration of Software Systems with Neural Interfaces (2024). Scientific Visualization 16.4: 11–24, DOI: 10.26583/sv.16.4.02
13. SI Chuprina . Using Data Fabric Architecture to Create Personalized Visual Analytics Systems in the Field of Digital Medicine (2023). Scientific Visualization 15.5: 50–63, DOI: 10.26583/sv.15.5.05
14. NA Bondareva . The Impact of Input Data Density on the Performance of Graphic Neural Networks (2024). Scientific Visualization 16.5: 109 - 119, DOI: 10.26583/sv.16.5.08
15. Morrone M., Exclusive: Popular chatbots amplify misinformation, 2025, - URL: https://www.axios.com/2025/09/04/popular-chatbots-amplify-misinformation?utm_source=Securitylab.ru
16. AI False Information Rate Nearly Doubles in One Year, 2025, - URL: <https://www.newsguardtech.com/ai-monitor/august-2025-ai-false-claim-monitor/>
17. Li S., Guo H., Tang X., Tang R., Hou L., Li R., Zhang R. Embedding Compression in Recommender Systems: A Survey (2024), arXiv:2408.02304
18. Vafa K., Chen JY, Rambachan A., Kleinberg J., Mullainathan S. Evaluating the World Model Implicit in a Generative Model (2024) <https://doi.org/10.48550/arXiv.2406.03689>
19. Torsten A. Enßlin . Information theory for fields (2018). Annalen der Physik 2019, vol. 531, issue 3, p. 1800127 DOI: 10.1002/andp.201800127
20. Bondarev A.E., Bondarenko A.V., Galaktionov V.A., Klyshinsky E.K. Visual analysis of cluster structures in multidimensional volumes of text information / Scientific visualization, Vol. 8, No. 3, 2016, pp. 1-24