УДК 004.658.2:025.4.03:004.8

# АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОГО ПОИСКА ИНФОРМАЦИИ В РАСПРЕДЕЛЕННЫХ БАЗАХ ДАННЫХ

Иович Е.А.

Белорусский государственный университет информатики и радиоэлектроники, Минск, Республика Беларусь, e.kozhemiako@bsuir.by

Аннотация: в статье рассматриваются современные алгоритмы интеллектуального поиска информации в распределенных базах данных. Основное внимание уделено индексированию, ранжированию, моделям машинного обучения для оценки релевантности, обработке распределённых запросов и методам оценки производительности поиска.

Ключевые слова: распределённые базы данных, индексирование, ранжирование, машинное обучение, релевантность, поисковые системы.

## I. ВВЕДЕНИЕ

Современные информационные системы работают с огромными объёмами данных, распределённых между множеством серверов и дата-центров. Эффективный поиск информации в таких условиях требует применения интеллектуальных алгоритмов, которые способны не только быстро находить документы, но и оценивать их релевантность. Проблема поиска в распределённых базах данных стала особенно актуальной с ростом облачных технологий, поисковых платформ и Big Data [1-13].

Методы индексирования. Индексирование является базовым этапом организации поиска. Наиболее распространённым подходом является обратный индекс, в котором каждому термину сопоставляется список документов. В распределённых системах применяются шаржированные индексы, где данные делятся по сегментам. Это позволяет выполнять параллельный поиск и уменьшать задержки. Также используются B-деревья, LSM-деревья и hash-based индексы. LSM-структуры особенно эффективны в системах с высокой интенсивностью записи. (рисунок 1).

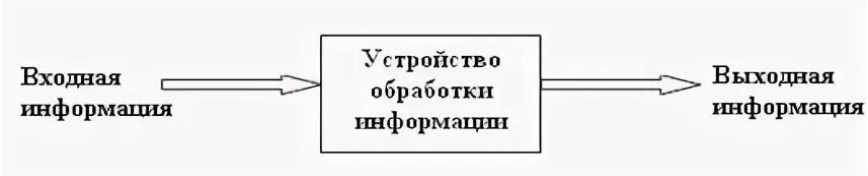



Рисунок 1. Общая схема процесса интеллектуального поиска информации

Представленная схема отражает базовую логику поискового процесса в распределённых системах: на этапе форматирования запрос преобразуется в структурированную форму, далее выполняется поиск по индексам, после чего результаты анализируются и при необходимости производится коррекция запроса. Такой итеративный подход позволяет повышать точность и полноту выдачи.

## II. АЛГОРИТМЫ РАНЖИРОВАНИЯ

После нахождения документов система должна определить порядок их выдачи. В современных системах применяются графовые методы ранжирования, такие как PageRank, а также поведенческие факторы: CTR, dwell time и история запросов. После нахождения документов система должна определить порядок их выдачи. Классическими методами являются TF-IDF и BM25. TF-IDF оценивает важность термина внутри документа, а BM25 учитывает длину документа и частоту термов.

ML-модели релевантности. Машинное обучение существенно улучшило качество поиска. Подход Learning to Rank позволяет обучать модели на размеченных данных. Среди популярных алгоритмов используются RankNet, LambdaMART и XGBoost. Нейросетевые архитектуры BERT и Transformer позволяют учитывать семантический контекст запроса, а не только совпадения ключевых слов. Это особенно важно для обработки естественного языка и длинных запросов. Архитектура подобной системы представлена на рисунке 2.

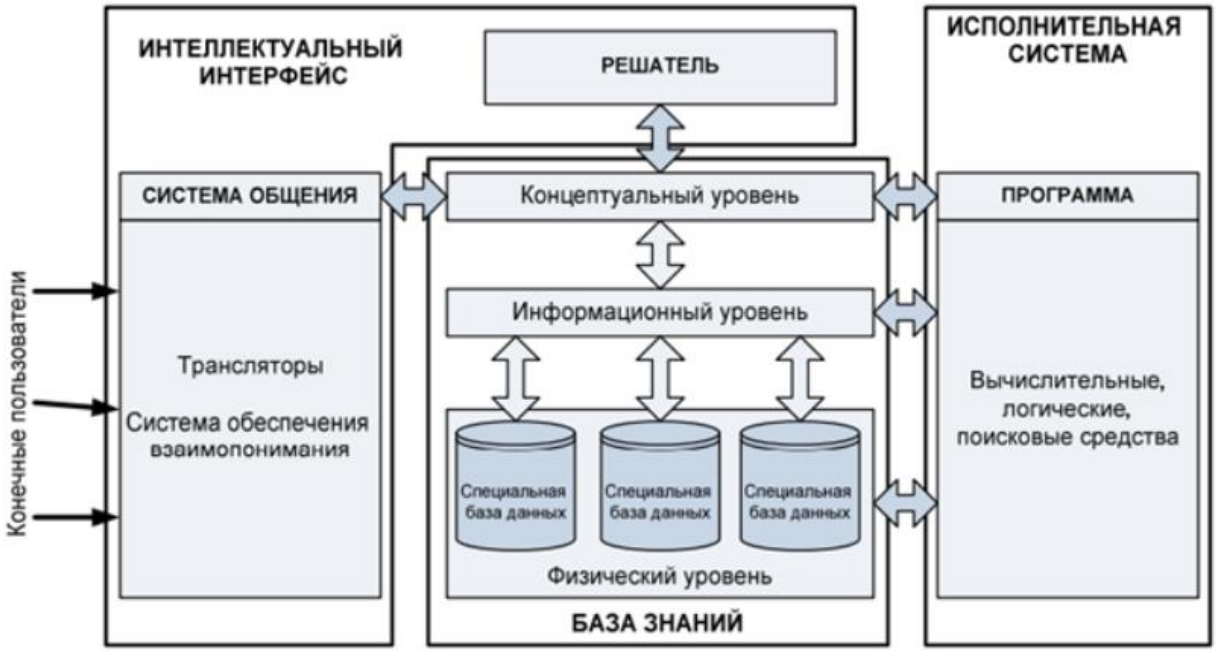



Рисунок 2. Архитектура интеллектуального анализа и выявления аномалий в распределённой системе

Обработка распределённых запросов. В распределённых базах данных запросы выполняются параллельно на нескольких узлах. Существует два основных подхода: scatter-gather и federated search. Scatter-gather предполагает отправку запроса всем узлам и последующую агрегацию результатов. Federated search объединяет результаты из разных независимых источников. Важной задачей является оптимизация сетевых задержек, балансировка нагрузки и минимизация времени агрегации.

Оценка скорости поиска. Производительность поисковых систем оценивается по метрикам latency, throughput и recall. Latency показывает среднее время ответа, throughput – число запросов в секунду. Для измерения качества поиска применяются Precision@K, Recall@K, MAP и NDCG. Оптимизация скорости достигается через кэширование, репликацию и асинхронную обработку запросов.

Перспективы развития. Будущее интеллектуального поиска связано с гибридными системами, объединяющими символический поиск и нейросетевые модели. Развитие векторных баз данных и approximate nearest neighbor (ANN) позволяет эффективно искать семантически похожие объекты. Интеграция LLM открывает возможности для диалогового поиска и генерации ответов на основе распределённых хранилищ данных.

## III. ЗАКЛЮЧЕНИЕ

Интеллектуальные алгоритмы поиска информации в распределённых базах данных являются основой современных поисковых платформ. Комбинация индексирования, алгоритмов ранжирования, машинного обучения и оптимизации распределённых запросов позволяет достигать высокой скорости и точности поиска. Дальнейшее развитие технологий будет связано с более глубоким использованием нейросетей и семантического анализа данных.

## БЛАГОДАРНОСТЬ

Работа выполнена на базе учреждения образования «Белорусский государственный университет информатики и радиоэлектроники». Автор выражает глубокую признательность руководству и коллективу центра за предоставленную инфраструктуру и техническую поддержку при проведении исследований.

## ЛИТЕРАТУРА

[1] Современные адаптивные системы в умном доме // Международный научно-исследовательский журнал. – 2024. – № 1(139). – DOI: 10.23670/IRJ.2024.139.162.

[2] Zhao, C. The Evolutionary Revolution of Smart Home Systems / C. Zhao // ACM Digital Library. – 2025. – DOI: 10.1145/3745238.3745421.

[3] Интеграция ИИ с системами умного дома / Elec.ru. – 2025. – [Электронный ресурс]. – Режим доступа: https://www.elec.ru/publications/it-ii/9345/.

[4] Энергоэффективность и экономия с AI: как нейросети помогают вашему дому тратить меньше / CDT5. – 2025. – [Электронный ресурс]. – Режим доступа: https://cdt5.ru.

[5] Архитектура системы «управления умным домом» / Science-engineering. – 2022. – [Электронный ресурс]. – Режим доступа: https://s.science-engineering.ru/pdf/2022/2/1388.pdf.

[6] Integration of AI and IoT for Smart Home Automation // International Journal of Scientific Research in Electrical and Computer Engineering. – 2024. – Vol. 11(5).

[7] Протоколы умного дома: от Zigbee до Matter и Thread / Neuracube. – 2025. – [Электронный ресурс]. – Режим доступа: https://neuracube.ru/blogs/blog/Zigbee-Matter-Thread.

[8] Протокол Matter для умного дома: как работает и что даёт / Lifehacker. – 2024. – [Электронный ресурс]. – Режим доступа: https://lifehacker.ru/protokol-matter/.

[9] Автоматизация умного дома с нейросетями / Future Media. – 2025. – [Электронный ресурс]. – Режим доступа: https://media.future-hub.io/stati/avtomatizatsiya-umnogo-doma-s-neyrosetyami.

[10] Зибко, И.В. Разработка системы управления «умным домом» / И.В. Зибко // БГУИР. – [Электронный ресурс]. – Режим доступа: https://libeldoc.bsuir.by/bitstream/123456789/42601/1/.

[11] RL-HEMS: Reinforcement Learning Based Home Energy Management System for HVAC Energy Optimization / ORNL // ASHRAE Transactions. – 2020. – Vol. 126. – PP. 421–429.

[12] Yu, L. Deep Reinforcement Learning for Smart Home Energy Management / L. Yu et al. // IEEE Internet of Things Journal. – 2019. – DOI: 10.1109/JIOT.2019.2957289.

[13] Дубровина, А.С. Анализ безопасности применения нейронных сетей в умном доме / А.С. Дубровина // МИД. – 2024. – С. 14–18.

## ALGORITHMS OF INTELLIGENT INFORMATION RETRIEVAL IN DISTRIBUTED DATABASES

E.A. Iovich

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus,
e.kozhemiako@bsuir.by

Abstract: the article discusses modern algorithms for intelligent information retrieval in distributed databases. The main attention is paid to indexing, ranking, machine learning models for relevance assessment, distributed query processing, and search performance evaluation methods.

Keywords: distributed databases, indexing, ranking, machine learning, relevance, search engines.