



---

## CHARACTERISTICS OF THE ACTIVITY OF METASEARCH SYSTEMS

Gulhayo Panjiyeva

Student Chirchik State Pedagogical University

panjiyevagulkhayo@gmail.com

---

### **Abstract:**

The article analyzes the role of metasearch systems in the information search process, their functional structure, algorithmic principles of operation, and their advantages over traditional search systems. The architecture of the systems, user interface, query routing, and result integration modules are considered from a technical point of view. The final part analyzes the future development prospects and scientific problems for modern metasearch systems based on artificial intelligence.

**Keywords:** Metasearch, LTR, syntactic analysis, semantic analysis, CTR, BERT, RoBERTa, cosine similarity, BM25, queries

### **ОСОБЕННОСТИ ДЕЯТЕЛЬНОСТИ МЕТАПОИСКОВЫХ СИСТЕМ**

Панджиева Гулхайё

Студентка Чирчикского государственного педагогического университета

panjiyevagulkhayo@gmail.com

### **АННОТАЦИЯ**

анализируется роль метапоисковых систем в процессе поиска информации, их функциональная структура, алгоритмические принципы работы и преимущества перед традиционными поисковыми системами. С технической точки зрения рассматриваются архитектура систем, пользовательский интерфейс, модули маршрутизации запросов и интеграции результатов. В заключительной части анализируются



---

дальнейшие перспективы развития и научные проблемы современных метапоисковых систем на основе искусственного интеллекта .

**Ключевые слова:** Метапоиск , LTR, синтаксический анализ, семантический анализ, CTR, BERT, RoBERTa, косинусное сходство, BM25, запросы

## **ВВЕДЕНИЕ**

В последние годы в глобальной интернет-среде произошел взрывной рост объема информации. Резкий рост количества страниц, статей, научных исследований, новостей и коммерческого контента в Интернете усложнил информационные потребности пользователей. Поэтому поисковые системы сегодня служат не только средством поиска нужной пользователю информации, но и ключевой платформой для навигации по информации, ее анализа и принятия решений.

Однако каждая поисковая система работает на основе собственных алгоритмов, механизмов индексации и моделей анализа. Это ограничивает результаты и приводит к однобокой информации для некоторых пользователей. В этом контексте все более актуальными становятся метапоисковые системы, то есть системы, которые объединяют сильные стороны нескольких поисковых систем и предоставляют более широкий, разнообразный и объективный результат на основе одного запроса.

*Теоретические основы метапоисковых систем : Метапоисковые системы впервые появились в 1990-х годах , и они отправляют запрос пользователя сразу нескольким независимым поисковым системам и отображают результаты в одном списке. Эти системы не индексируют веб-страницы напрямую , а полагаются на существующие поисковые системы .*

Подход метапоиска основан на следующих принципах:

*Делегирование* : задача сбора данных делегируется внешним поисковым системам.

*Агрегация* : результаты, полученные из разных систем, суммируются.

*Переранжирование* : наиболее релевантные результаты ранжируются по приоритету путем переранжирования.



---

*Нормализация* : результаты приводятся во взаимно совместимую форму (например, форматирование URL, единица оценки).

### **Архитектура и компоненты**

Метапоисковые системы состоят из следующих основных модулей:

*Пользовательский интерфейс (UI)*. Этот модуль напрямую взаимодействует с пользователем. Он содержит строку поиска, фильтры, список результатов, механизмы рейтинга и рекомендаций. UI создается с использованием современных веб-технологий (Vue.js, React.js).

*Модуль маршрутизации запросов*. Этот модуль отправляет запрос пользователя в требуемом формате в несколько поисковых систем (например, Google Custom Search API, Bing API). На этом этапе решаются вопросы аутентификации, лимита и формата.

*Модуль сбора и интеграции результатов*. Результаты, возвращаемые по запросу, собираются в формате JSON или XML, нормализуются (заголовок, ссылка, краткое описание) и объединяются в общий список. На этом этапе используются функции для обнаружения дубликатов, пересчета рейтинга и адаптации к профилю пользователя.

### **Используемые алгоритмы и модели**

Главным преимуществом метапоисковых систем является обработка, развитие и оптимальная сортировка данных, полученных из источников. В процессе работы широко используются современные алгоритмы и модели обучения.

- *BM25 / TF-IDF – Оценка релевантности на основе текста*

Эти две модели основаны на важных словах для текстов (увеличение сходства).

- **TF-IDF** (Term Frequency-Inverse Document Frequency) - оценивает важность каждого слова: сколько раз слово встречается в тексте (TF) и как редко слово встречается в других документах (IDF). Эта модель служит базовым шагом для общих устройств.

- **BM25** — это улучшенная версия модели TF-IDF, которая более глубоко анализирует взаимосвязь между словами. Она учитывает частоту и частоту появления документа, что помогает предоставлять более точные



---

разрешения в реальных ситуациях. BM25 особенно эффективен при сравнении разных источников.

- *Косинусное сходство – измерение сходства результатов.*

**Косинусное сходство** - измеряет внешний угол между двумя текстовыми векторами и определяет сходство файлов. Чем меньше угол (ближе к  $0^\circ$ ), тем более похожи тексты. Этот алгоритм особенно важен в пользовательских запросах и семантической помощи почти самому себе. Перекрестные ссылки на тексты в векторном пространстве позволяют сопоставлять их.

- *Обучение ранжированию (LTR) – Машинное обучение для ранжирования высокого уровня*

**LTR** — это набор правил, которые регулируют программное обеспечение путем проверки безопасности прав пользователей.

- В этом методе модель машинного обучения строится *на основе использования времени* (клики, затраченное время, прокрутка и т. д.).

- Среди моделей LTR популярны такие алгоритмы, как *RankNet*, *LambdaRank* и *LambdaMART*, которые основаны на градиентном бустинге, нейронных сетях и других методах машинного обучения.

- Метапоисковые системы LTR оказывают значительную помощь *пользователям в поиске нужного программного обеспечения*, особенно при интеграции приложений из нескольких источников.

- *BERT, RoBERTa – Модели Transformer для семантического анализа* пытаются углубить запрос пользователя не только с помощью синтаксического, но и **семантического анализа**.

- **BEPT** (Представления двунаправленного кодировщика от Transformers) - определяет значение каждого слова на основе контекста. BERT анализирует текст, беря контекстный материал с обеих сторон (слева и справа), определяя точность этой основы.

- **RoBERTa** — это улучшенная версия BERT, обученная на большем количестве данных и с улучшенным обучением. RoBERTa обеспечивает высокоточные результаты, особенно для сложных запросов или неоднозначных вопросов.



- 
- Эти модели используются в системах, которые *понимают запросы* , сортируют документы , отвечают на вопросы и выполняют поиск отрывков.
  - *Статистика кликабельности (CTR) – для ранжирования лечения заболеваний*  
**показатель кликабельности )**
    - *выберет наиболее релевантные результаты для пользователей* на основе CTR. изучает информацию и на ее основе дает оценки последующим пользователям.
    - Эта информация *используется для понимания намерений пользователя и персонализации.* для.
    - *Скорость прокрутки* вместе со статистикой CTR , время, проведенное на странице , нажатия кнопок такие проверки также применяются к планированию сортировки.

#### **Области применения**

*В научных исследованиях* метапоисковые системы ищут статьи или источники из нескольких научных баз данных в единой структуре. Например, Google Scholar и Semantic Scholar могут использоваться параллельно для поиска необходимой научной помощи. Это экономит время научных исследователей и обеспечивает комплексный анализ.

*Коммерческие и торговые* метапоиски очень помогают в сравнении цен. По сути, когда вы вводите название продукта, система находит его в разных магазинах, сравнивает цены и доступные варианты, а затем разделяет его. Примерами могут служить PriceRunner или Shopzilla.

*В сфере туристических услуг* метапоисковые системы собирают информацию с различных сайтов по продаже авиабилетов или аренды автомобилей и предлагают пользователям наиболее удобные и доступные варианты. Примерами являются такие платформы, как Skyscanner или Kayak.

*Медицина* - Объединение метапоиска с несколькими инструментами анализа симптомов (инструментами анализа симптомов) помогает точно определить симптомы пациента, вызывающие болезнь или заболевание. Это особенно полезно для последующих действий.



---

## **ЗАКЛЮЧЕНИЕ**

*Системы поиска метаданных* — это удобный и мощный альтернативный способ поиска информации в современной информационной среде. Эти системы собирают ресурсы из различных источников и представляют их пользователю в одном месте. Этот ярлык открывает путь к более быстрой разработке процесса поиска информации, сравнению большего количества программ и более широкому анализу.

*Преимущества* включают экономию времени, унифицированное использование, простую установку и простоту использования. Недостатком является то, что эти системы технически сложны и зависят от API-сервисов многих организаций. Иногда это может приводить к задержкам, расходам или потерям в производительности. , адаптируясь к хранению продуктов, используемых многими людьми, т. е. не давая персонализированных рекомендаций.

*В будущем есть несколько важных тенденций* . В частности, благодаря внедрению многоязычных систем пользователь может отправлять запрос на своем языке и получать веб-сайты на других языках. Создание семантических технологических систем на основе моделей искусственного интеллекта, особенно таких трансформеров, как BERT и GPT, сделает процесс поиска более интеллектуальным и интуитивно понятным. Это поможет использовать его для офлайн-использования. Персонализация — это возможность анализировать историю поиска пользователя и предоставлять индивидуальные рекомендации, подходящие именно ему.

*В целом* , метапоисковые системы являются полезным инструментом для быстрого, надежного и масштабируемого поиска, помогая повысить эффективность в *различных* областях — от научных исследований до коммерции и путешествий. По мере того, как они становятся все более интеллектуальными, эти системы становятся все более и более ориентированными на пользователя.



---

## **Литература**

1. Крапивин Ю. Метод глубокого обучения в контексте информационного поиска для решения задач автоматической генерации естественного языка и автоматизированной генерации текста // Труды конференции «Восток-Запад» по интеллектуальным системам, Международная конференция, 16–18 сентября 2017 г. Белорусский государственный университет информатики и радиоэлектроники. - Минск, 2017. - С. 299-302.
2. Мэннинг, К., Рагхаван, П., Шютце, Х. Введение в информационный поиск. Cambridge University Press, 2008.
3. Баеза-Йейтс, Р., Рибейро-Нето, Б. Современный информационный поиск. Addison Wesley, 2011.
4. Google Developers. Документация по пользовательскому поиску JSON API – <https://developers.google.com/custom-search>
5. Документация API поиска Microsoft Bing – <https://learn.microsoft.com/en-us/bing/search-apis/>