

Эволюция поисковых систем в электронной коммерции: от статических каталогов к гибридным и векторным архитектурам

П.С. Астахов, А.Г. Белокрылов, Ф.В. Краснов

Аннотация — Статья посвящена эволюции поисковых систем в электронной коммерции с акцентом на сегмент DIY (Do-It-Yourself). Рассматривается переход от статических каталогов и SEO-оптимизации к современным гибридным архитектурам, объединяющим методы Learning-to-Rank, векторный поиск и мультимодальные модели. Особое внимание уделено инженерным вызовам, связанным с масштабируемостью, поддержанием актуальности данных и интеграцией машинного обучения в промышленные решения. Новизна исследования заключается в формулировке поэтапной стратегии трансформации поисковых систем, адаптированной к специфике DIY-сегмента и опирающейся на анализ практик ведущих ритейлеров (Amazon, Walmart, Wayfair, Home Depot, Alibaba, Ozon, IKEA, Wildberries). Дополнительно выдвигаются исследовательские гипотезы и предлагается экспериментальный протокол, позволяющий связать теоретическую модель с эмпирической проверкой и объективной оценкой качества поиска. Таким образом, статья сочетает исторический и прикладной анализ, формирует целостную картину развития поисковых технологий и определяет перспективные направления их применения в условиях растущей сложности пользовательских запросов и динамики ассортимента.

Ключевые слова — Гибридный поиск, Learning-to-Rank, Машинное обучение, Трансформация, Информационный поиск, Архитектура поисковых систем

I. ВВЕДЕНИЕ

Поиск является центральным элементом взаимодействия пользователей с платформами электронной коммерции, особенно в сегменте DIY, где запросы часто носят узкоспециализированный характер, связанный с конкретными задачами, такими как подбор инструментов, крепежных элементов или расходных материалов [3, 11]. Качество поисковой системы напрямую определяет конверсию, уровень удовлетворенности пользователей и ключевые бизнес-показатели, включая средний чек и долгосрочную ценность клиента (LTV). В отличие от сегментов одежды или электроники, DIY характеризуется высокой долей технических запросов, требующих точной интерпретации и обработки сложных атрибутов, таких

как размеры, материалы и стандарты совместимости [9]. Это делает разработку поисковых систем для DIY-сегмента одновременно инженерной и научной задачей, требующей междисциплинарного подхода.

Цель настоящего исследования — проанализировать эволюцию поисковых систем в электронной коммерции, уделяя особое внимание DIY-сегменту, и предложить стратегию трансформации от традиционных статических каталогов к гибридным архитектурам, основанным на машинном обучении и векторном поиске. Новизна заключается в акценте на поэтапной трансформации, минимизирующей риски и адаптированной к специфическим требованиям DIY, таким как обработка разговорных запросов и интеграция мультимодальных данных. Статья структурирована следующим образом: исторический обзор эволюции поисковых систем, анализ современных архитектур, инфраструктурные вызовы, стратегия трансформации, подходы к валидации и экспериментальный протокол. Этот анализ закладывает основу для понимания исторического контекста, который рассматривается далее с прицелом на выполнение трансформации.

В контексте этих целей важно отметить, что эволюция поисковых систем не только отражает технологические инновации, но и напрямую коррелирует с изменениями в поведении потребителей, где DIY-запросы эволюционировали от простых ключевых слов к сложным семантическим конструкциям, требующим интеграции ИИ. Такой подход позволяет связать исторический анализ с практическими рекомендациями, подчеркивая необходимость междисциплинарного синтеза [1, 3].

II. ИСТОРИЧЕСКИЙ ОБЗОР

Эволюция поисковых систем в электронной коммерции отражает технологический прогресс и изменения в пользовательских ожиданиях. На ранних этапах, в 1990-х годах, первые онлайн-магазины, такие как Amazon и eBay, полагались на статические каталоги с иерархической структурой категорий [5, 16]. Поиск сводился к навигации по дереву категорий, что обеспечивало приемлемую производительность при ограниченном ассортименте, но становилось неэффективным с увеличением числа SKU. Исследование Amazon подчеркивает, что такие системы плохо справлялись с обработкой текстовых запросов,

Статья получена 1 ноября 2025.

П.С. Астахов, pavel.astakhov@vseinstrumenti.ru,

А.Г. Белокрылов, aleksejj.belokrylov@vitech.team,

Ф.В. Краснов, fkrasnov2@yandex.ru, <http://orcid.org/0000-0002-9881-7371>.

особенно в условиях роста ассортимента, что стимулировало переход к более динамичным подходам [6]. Именно на этом этапе стала очевидной необходимость учета пользовательских интенгов, что заложило основу для последующих инноваций.

Этот последовательный переход отражает не только технологический прогресс, но и адаптацию к изменяющимся потребностям пользователей, где ключевую роль сыграли компании вроде Amazon и Alibaba, внедрявшие инновации для повышения релевантности поиска [5, 24]. В контексте DIY-сегмента такая эволюция была особенно значима, поскольку технические запросы требовали большего внимания к семантике и атрибутам товаров, что привело к более медленному, но целенаправленному развитию.

В начале 2000-х годов полнотекстовый поиск стал ответом на потребность в более гибкой обработке запросов. Компании, такие как Alibaba и Ozon, внедрили системы, основанные на ключевых словах, что позволило пользователям находить товары по их описаниям и названиям [24, 16]. Исследование эволюции Taobao, принадлежащего Alibaba, показывает, что внедрение полнотекстового поиска в связке с Alipay в 2003 году не только решило проблему доверия в транзакциях, но и увеличило рыночную долю до 50% в Китае, подчеркивая важность поисковых технологий для конкурентоспособности [25]. Аналогично, Ozon в России адаптировал полнотекстовый поиск для обработки локализованных запросов, что позволило платформе эффективно конкурировать в сегменте электроники и DIY [16]. Однако полнотекстовый поиск требовал более точного учета пользовательского поведения, что привело к следующему этапу.

Сравнивая эти этапы, можно отметить, что если в 1990-х статические каталоги обеспечивали базовую навигацию, то внедрение полнотекстового поиска в 2000-х ознаменовало сдвиг к пользователь-центричным системам, где ключевую роль сыграли алгоритмы, подобные BM25, позволяющие учитывать частотность терминов и длину документов [13]. Этот анализ подчеркивает, как постепенная интеграция поведенческих данных заложила основу для LtR-моделей, минимизируя ограничения предыдущих подходов и повышая общую эффективность в динамичных рынках, таких как DIY [6, 24].

Использование *implicit feedback*, начавшееся в середине 2000-х, позволило интегрировать поведенческие данные, такие как клики, просмотры и покупки, для повышения релевантности выдачи. eBay в 2021 году представила двухбашенную модель для персонализированных эмбедингов, которая учитывала пользовательские взаимодействия, решая проблему холодного старта для новых товаров и обеспечивая охват до 90% просмотров страниц [26]. Похожий подход применила Wildberries, крупнейшая российская e-commerce платформа, которая использовала

поведенческие данные для оптимизации выдачи в категориях с высокой конкуренцией, таких как DIY и бытовая техника [17]. Подобные системы продемонстрировали переход от статических к динамическим подходам, обозначив важный рубеж в развитии поиска.

Внедрение моделей Learning-to-Rank (LtR) в конце 2000-х годов ознаменовало новый этап, позволив учитывать множество признаков, включая текстовые, поведенческие и коммерческие. Исследование экосистемы Alibaba показало, как LtR-модели, интегрированные с данными Taobao и Alipay, обеспечили конкурентное преимущество в C2C-сегменте, позволив платформе расшириться на 13 отраслей [24]. Home Depot в 2021 году представила end-to-end решение для Named Entity Recognition (NER) в поиске, где фреймворк TripleLearn повысил F1-score с 69.5 до 93.3, улучшив конверсию за счет точного распознавания сущностей в запросах [19]. Эти достижения подчеркивают роль машинного обучения в обработке сложных запросов, особенно в DIY-сегменте. Современный этап, начавшийся в 2010-х годах, характеризуется внедрением гибридных архитектур, которые объединяют традиционные методы (булевый поиск, BM25) с векторным поиском, обеспечивая обработку как семантических, так и long-tail запросов. Walmart в 2024 году разработал семантическую систему ретривала для рекламы с использованием LLM, обученных на нескольких доменах знаний, что улучшило nDCG на 16% и сократило нерелевантные объявления [27]. Wayfair в 2023 году выпустила датасет WANDS с 480 запросами и 233,000 аннотациями, созданный для оценки релевантности поиска, что способствовало развитию моделей и улучшению метрик, таких как nDCG [22]. Эти исследования демонстрируют, как гибридные системы решают проблему обработки сложных запросов, что особенно актуально для DIY.

Причины изменения подходов к индексированию связаны с ростом объемов данных, увеличением разнообразия запросов и необходимостью обработки long-tail запросов. Статические каталоги не справлялись с динамическими изменениями ассортимента, что привело к внедрению полнотекстового индексирования на основе *inverted indexes*, таких как Lucene, используемого в Elasticsearch [13]. Однако полнотекстовый поиск оказался недостаточным для семантических запросов, что стимулировало переход к векторному индексированию с использованием ANN-алгоритмов, таких как HNSW и PQ [4]. Исследование Amazon показало, что переход к векторным индексам позволил сократить latency на 30% для long-tail запросов [6]. Аналогично, Wildberries внедрила гибридное индексирование для обработки локализованных DIY-запросов, что улучшило релевантность на 15% [17]. Эти изменения были обусловлены необходимостью баланса между скоростью, точностью и масштабируемостью, что подводит к анализу современных архитектур.

Эволюция от статических каталогов к гибридным системам демонстрирует не только технологический прогресс, но и изменение парадигмы восприятия поиска – от простого инструмента навигации к интеллектуальному посреднику между пользователем и обширным каталогом. Этот переход напрямую подводит нас к рассмотрению архитектуры современных систем, которые интегрируют различные методы для обеспечения высокой точности и эффективности.

Таким образом, историческая траектория показывает не только эволюцию технологий, но и постепенное усложнение пользовательских ожиданий. Эти изменения создают основу для понимания архитектуры современных поисковых систем, которая объединяет элементы всех предшествующих подходов.

III. АРХИТЕКТУРА СОВРЕМЕННЫХ ПОИСКОВЫХ СИСТЕМ

Современные поисковые системы можно рассматривать как многослойные конструкции, в которых простые механизмы фильтрации и полнотекстового поиска сочетаются с семантическими и обучаемыми компонентами. Такой синтез обусловлен как ростом ассортимента, так и необходимостью точного захвата пользовательских интенгов.

Ключевые элементы поисковые системы:

- **Candidate Generation:** Формирование пула релевантных товаров, обеспечивающее полноту поиска [10].
- **Pre-ranking:** Предварительное ранжирование для ускорения обработки популярных запросов, минимизирующее вычислительные затраты.
- **Онлайн-ранжирование:** Использование LtR-моделей для учета текстовых, поведенческих и коммерческих признаков, таких как маржинальность и популярность [1].
- **Семантический поиск:** Векторные представления и ANN-индексы для обработки long-tail и разговорных запросов [4].
- **Session-aware модели:** Моделирование последовательностей действий пользователей для повышения релевантности в многошаговых сценариях [12].

Анализируя эти элементы, стоит подчеркнуть, что в гибридных архитектурах ключевым фактором успеха становится баланс между вычислительной эффективностью и семантической точностью, как демонстрирует пример Walmart, где LLM-модели

интегрированы для семантического ретривала, улучшая nDCG на 16% [27]. Такой синтез не только решает проблемы масштаба, но и адаптирует систему к специфике DIY, где мультимодальные данные (текст + изображения) позволяют лучше захватывать интенды пользователей [11, 22].

Гибридные архитектуры, комбинирующие булевый поиск, BM25 и векторные методы, обеспечивают оптимальный баланс между точностью и полнотой [10]. В DIY-сегменте, где запросы часто носят разговорный характер (например, «чем заменить треснувший патрон в дрели»), векторный поиск позволяет учитывать семантику запроса, а LtR-модели обеспечивают приоритизацию релевантных товаров. Например, Home Depot использует гибридные системы для обработки запросов, связанных с совместимостью инструментов и аксессуаров, что улучшает пользовательский опыт [19].

Углубленный анализ гибридного поиска показывает, что его эффективность в DIY-сегменте обусловлена не только комбинацией методов, но и адаптивным использованием алгоритмов слияния результатов, таких как RRF или learning-based fusion, которые позволяют динамически взвешивать вклады от keyword-matching и векторного ретривала в зависимости от типа запроса [10, 13]. Например, для технических запросов с точными атрибутами (размеры, материалы) преобладает булевый фильтр, обеспечивая precision, в то время как для разговорных или long-tail запросов (например, «альтернатива сломанному патрону в дрели») векторный поиск повышает recall, захватывая семантические синонимы и контекст [4, 22]. Это позволяет гибридным системам достигать улучшения метрик на 10–20% по сравнению с монометодами, как показано в кейсах Walmart и Home Depot, где интеграция LLM с традиционными индексами сократила нерелевантные результаты и повысила конверсию [19, 27]. Однако такой подход требует тщательной калибровки гиперпараметров, чтобы избежать переобучения на доменных данных, и учета trade-off между вычислительными затратами и качеством, особенно в условиях реал-тайм обработки миллионов SKU.

Для иллюстрации преимуществ гибридного поиска в практике рассмотрим кейсы Amazon, лидера в e-commerce, включая DIY-категории. Внедрение PECOS (Prediction for Enormous and Correlated Output Spaces Framework) для product retrieval и text autocompletion позволило Amazon организовать пространство продуктов в семантические кластеры, сократив время обработки запросов и повысив recall на 15–20% для long-tail запросов, таких как «альтернативы для ремонта дрели в домашних условиях». Это решение, интегрированное с ML-based matching, минимизировало нерелевантные результаты, как показано в анализе Amazon Search Ranking с использованием nDCG-метрики: модель на основе gradient boosting достигла

улучшения nDCG на 12% по сравнению с базовыми алгоритмами, напрямую влияя на конверсию в категориях инструментов и материалов. Аналогично, в AWS Retail Demo Store персонализированный ranking с Amazon Personalize-Ranking recipe повысил conversion rates на 18% за счёт реал-тайм адаптации к пользовательским предпочтениям, демонстрируя, как такие метрики коррелируют с ростом LTV в динамичных сегментах вроде DIY. Эти примеры подчёркивают, что гибридные системы не только решают теоретические вызовы, но и приносят measurable бизнес-результаты, подтверждённые A/B-тестами в production-средах.

Аналогичный подход демонстрирует Home Depot, ведущий DIY-ритейлер, интегрировавший элементы Google Shopping в свою поисковую систему для обработки запросов по совместимости инструментов и аксессуаров. Согласно данным, внедрение гибридных кампаний, сочетающих товарные объявления с локальным инвентарем, привело к росту трафика в физические магазины на 15% и онлайн-конверсий на 22%, с улучшением метрики nDCG за счёт семантического анализа контекста запросов, таких как «альтернативы для ремонта дрели» [30, 31]. Этот кейс подчеркивает инфраструктурные аспекты трансформации: оптимизация миллионов SKU через автоматизированные фиды снизила ошибки на 40%, но потребовала значительных инвестиций в ML-модели для реал-тайм обновлений, иллюстрируя brown field подход с gradual миграцией от legacy-систем к интеллектуальным. В сравнении с меньшими игроками рынка, здесь акцент на масштабе подтверждает эффективность гибридных архитектур в повышении релевантности и бизнес-метрик в динамичных сегментах e-commerce.

В контексте DIY-сегмента аналогичные принципы иллюстрирует IKEA, ритейлер мебели и товаров для дома, где оптимизация поисковой системы в e-commerce фокусируется на предиктивном поиске, фильтрах и персонализированных рекомендациях для улучшения навигации по категориям самосборки и ремонта. Внедрение этих функций, включая автозаполнение с учетом синонимов и организованных фильтров по атрибутам, способствовало значительному росту онлайн-продаж (по оценкам, на 20%) и косвенному повышению конверсии (на 12%) за счёт упрощенного процесса поиска и рекомендаций. Этот кейс подчеркивает роль данных аналитики в персонализации, где алгоритмы учитывают предпочтения пользователей в DIY-запросах (например, «альтернативы для сборки шкафа»), снижая время поиска и повышая удовлетворенность, с ростом показателей лояльности на 15% в связанных категориях. В сравнении с крупными платформами, как Amazon, здесь акцент на интеграции с мобильными приложениями подтверждает эффективность гибридных подходов в минимизации ошибок и повышении релевантности результатов для

динамичных ассортиментов, требуя, однако, тщательного аудита UX для устранения неточностей в фильтрах.

Однако внедрение таких архитектур сопряжено с вычислительными издержками, особенно при использовании GPU для генерации эмбеддингов. CUDA, как платформа NVIDIA для параллельных вычислений, эффективна для задач ANN-поиска, но её зависимость от специализированного оборудования ограничивает гибкость масштабирования [4]. В распределенных системах координация нескольких GPU на разных нодах увеличивает задержки из-за передачи данных, что снижает производительность при высоких нагрузках. Ограниченная память GPU (16–80 ГБ) создает сложности при работе с крупными векторными индексами, характерными для DIY-сегмента с миллионами SKU [7]. Шардинг индекса по нескольким GPU добавляет сложность в синхронизацию, что может привести к несогласованности данных и задержкам в обновлениях (freshness). Проблемы совместимости CUDA в гетерогенных CPU/GPU-кластерах также могут вызывать сбои, что особенно критично для e-commerce, где требуется высокая доступность [13]. Перечисленные архитектурные особенности подчеркивают, что сама структура поисковой системы задаёт высокие требования к инфраструктуре.

Таким образом, необходимость масштабируемости, отказоустойчивости и поддержки гибридного индексирования порождает специфические инженерные вызовы, которые необходимо рассмотреть при реализации систем в условиях реального рынка.

IV. ИНФРАСТРУКТУРНЫЕ ВЫЗОВЫ

Реализация современных поисковых систем в электронной коммерции сопряжена с рядом инженерных проблем, которые вытекают из архитектурных особенностей и требований DIY-сегмента.

Ключевые вызовы включают в себя:

- **Баланс полноты и латентности:** Оптимизация скорости ответа при сохранении качества поиска требует сложных алгоритмов и инфраструктуры [4].
- **Актуальность данных (freshness):** Динамически изменяющийся ассортимент DIY-сегмента (например, изменения в наличии, цене, технических атрибутах или рейтингах) требует реал-тайм обновлений, что критически важно для точности LtR-моделей и конверсии.
- **Масштабирование ANN-поиска:** Алгоритмы HNSW и PQ обеспечивают компромисс между скоростью и качеством, но требуют значительных вычислительных ресурсов [4, 7].

Важно подчеркнуть, что данные вызовы носят не только

технический характер: они напрямую связаны с бизнес-показателями, поскольку задержки и ошибки в поиске приводят к снижению конверсии и росту числа отказов.

Amazon использует распределенные системы для обработки логов и кэширования, минимизируя задержки и поддерживая высокую производительность [6]. В отличие от этого, DIY-ритейлеры, такие как Home Depot, часто полагаются на стандартные возможности Elasticsearch, что ограничивает масштабирование сложных ML-моделей [8]. Ozon и Wildberries адаптировали облачные решения, такие как AWS OpenSearch, для обработки пиковых нагрузок, что позволило сократить latency на 20% для локализованных запросов [16, 17]. Проблема freshness особенно критична для DIY, где доступность товаров напрямую влияет на пользовательский опыт.

Дополнительные вызовы связаны с интеграцией с legacy-системами, которые часто имеют неструктурированные данные или устаревшие форматы, затрудняя построение обратных индексов [13]. Высокая стоимость GPU-кластеров для векторного поиска может достигать тысяч долларов в месяц, что требует оптимизации через облачные сервисы [14]. Энергопотребление GPU-кластеров также становится фактором, особенно в контексте устойчивого развития. Безопасность данных в распределенных системах требует шифрования и контроля доступа, чтобы предотвратить утечки, особенно в DIY, где обрабатываются персонализированные данные. Мониторинг производительности с использованием инструментов, таких как Prometheus или Kibana, необходим для выявления узких мест. Эти инфраструктурные аспекты напрямую влияют на стратегию трансформации, описанную в следующем разделе.

A. Обратные индексы (inverted indexes, Lucene/BM25)

Преимущества: высокая скорость, экономное использование памяти, зрелая экосистема (Elasticsearch/OpenSearch), поддержка фильтров и агрегаций.

Ограничения: слабая семантическая релевантность, трудности с перефразировками и разговорной лексикой.

B. Векторные индексы (ANN: HNSW, IVF+PQ, Faiss)

Преимущества: высокая семантическая полнота, устойчивость к перефразам, возможность объединять мультимодальные признаки.

Ограничения: сложность обновлений (вставки/удаления), потребность в памяти/CPU/GPU, approximate nature (погрешности), необходимость подбора гиперпараметров (efConstruction/efSearch для HNSW, nlist/codebook для IVF+PQ).

C. Гибридное индексирование

Комбинация inverted + vector index — наиболее

практичный путь: keyword-matching гарантирует точную фильтрацию (например, «размер», «материал», «совместимость»), а vector-retrieval увеличивает recall и захватывает разговорные/семантические запросы. Алгоритмы объединения результатов: RRF (Reciprocal Rank Fusion), learning-based fusion (фили из обеих выдач подаются в LtR), ранжирование по ансамблю моделей. Гибридные системы важны для DIY, где часты «точные» технические фильтры и одновременно требуются семантические соответствия.

D. Freshness и инкрементальные обновления

Проблема актуальности индексов критична для e-commerce. Существуют следующие практики решения:

- Delta indices / nearline updates: периодическое обновление основного индекса и отдельный «delta» индекс для свежих изменений; запрос объединяет оба.
- Stream processing & change data capture: события inventory/price/availability обновляются в реальном времени и попадают в очередь для быстрых обновлений эмбедингов/метаданных.
- TTL и soft-deletes: для облегчения удалений и отметки артефактов.

Каждый из подходов имеет свою цену: дельта-индекс увеличивает латентность комплексного запроса; Обновления в реальном времени требуют быстрой обработки эмбедингов (возможно на графическом процессоре) и оркестрации.

В контексте этих вызовов анализ показывает, что гибридное индексирование не только минимизирует ограничения отдельных подходов, но и позволяет интегрировать реал-тайм обновления через stream processing, как в случае с Alibaba, где такая стратегия обеспечила конкурентное преимущество в C2C-сегменте [24]. Это подтверждает необходимость баланса между freshness и латентностью, где delta-индексы служат промежуточным решением для DIY-ритейлеров с динамичным ассортиментом [7, 13].

Наличие перечисленных вызовов демонстрирует, что внедрение гибридного поиска невозможно без учёта инженерных ограничений и выстраивания стратегии пошаговой трансформации. Следующий раздел раскрывает эти аспекты.

V. ИНЖЕНЕРНЫЕ СЛОЖНОСТИ И СТРАТЕГИЯ ТРАНСФОРМАЦИИ ПОИСКОВЫХ СИСТЕМ

Новизной данного исследования является разработка поэтапной стратегии трансформации поисковых систем от традиционных статических каталогов к гибридным ML- и векторным архитектурам, адаптированным для DIY-сегмента. Эта стратегия учитывает инфраструктурные ограничения и минимизирует риски,

такие как сбои в производительности и потеря данных, обеспечивая преобразование каталога и создание высокопроизводительных систем обработки запросов [10, 12].

При реализации трансформации следует учитывать инженерные сложности:

- Проблема равноправных кандидатов: Булевый поиск считает все товары, соответствующие запросу, равноправными, что снижает точность для сложных DIY-запросов, требующих дифференциации по атрибутам [4].
- Формирование данных из бизнес-систем: Неструктурированные данные из ERP и CRM требуют сложного ETL-процесса, увеличивая затраты на обработку и хранение [13, 14, 15].
- Интеграция ML-моделей: Баланс между точностью эмбедингов и скоростью, а также обработка ошибок в векторизации [11].
- Масштабируемость: Обработка миллионов запросов с низкой латентностью требует оптимизации batch-процессинга [13].
- Адаптация к DIY: Технические характеристики товаров усложняют векторизацию и требуют мультимодальных моделей.
- Подготовка данных для LtR: Сбор и разметка данных для моделей ранжирования являются затратными, особенно в DIY, где требуется учет технических атрибутов [10, 15].

Перечисленные сложности можно условно разделить на три группы: связанные с качеством данных, с интеграцией моделей и с масштабируемостью инфраструктуры.

При разработке стратегии трансформации критически важно учитывать текущее состояние поисковой системы в компании как *brown field project*, то есть существующую инфраструктуру с *legacy*-системами, накопленными данными и интеграциями, которые нельзя игнорировать или полностью перестраивать без рисков для бизнес-процессов [13, 14]. В отличие от *green field* подхода, где все строится с нуля, *brown field* требует инвентаризации существующих компонентов (например, устаревших статических каталогов или базовых полнотекстовых индексов), оценки их совместимости с новыми ML-моделями и поэтапной миграции данных, чтобы минимизировать *downtime* и потерю функциональности. Для DIY-ритейлеров это означает анализ зависимостей от ERP-систем, обработку неструктурированных атрибутов товаров и постепенную интеграцию векторных индексов параллельно с существующими, с использованием A/B-тестов для валидации на подмножествах трафика [10, 15]. Такой

подход снижает риски, связанные с прерыванием сервиса, и позволяет адаптировать трансформацию под специфику компании, повышая общую эффективность внедрения.

Рассмотренные инженерные сложности показывают, что внедрение новых поисковых технологий невозможно осуществить одномоментно без значительных рисков для стабильности системы. Именно поэтому компании прибегают к постепенной трансформации, при которой каждая стадия сопровождается контролем метрик и ограничением потенциальных сбоев. Такая стратегия позволяет не только поэтапно адаптировать инфраструктуру и данные, но и последовательно обучать команды работе с новыми инструментами. В совокупности это формирует основу для поэтапной стратегии трансформации, которая подробно представлена далее. Поэтапная стратегия трансформации:

1. Оценка текущей системы и планирование (1–2 месяца): Аудит инфраструктуры с использованием Apache JMeter, анализ каталога, выбор инструментов (Elasticsearch/OpenSearch, модели эмбедингов) [13, 14].
2. Подготовка данных и каталога (3–5 месяцев): Очистка данных, нормализация, генерация эмбедингов, разметка для LtR-моделей [11, 15].
3. Внедрение гибридного поиска (2–3 месяца): Настройка keyword-based и векторного поиска, интеграция RRF, тестирование прототипа [13, 14].
4. Оптимизация и тестирование (2–4 месяца): Тюнинг моделей, A/B-тесты, интеграция behavioral analytics [9].
5. Масштабирование и мониторинг (1–2 месяца): Развертывание в production, использование облачных сервисов, настройка дашбордов [10].

Эта стратегия обеспечивает плавный переход, минимизируя риски и повышая конверсию в DIY, что связано с необходимостью качественного сбора данных, обсуждаемого далее.

Однако даже самая продуманная архитектура не будет успешной без качественных данных. Поэтому логичным шагом становится рассмотрение задач feature engineering и подготовки данных, обеспечивающих эффективность гибридных моделей.

VI. СБОР ДАННЫХ И FEATURE ENGINEERING

Подготовка данных становится связующим звеном между архитектурой и качеством итоговой выдачи. В условиях DIY-сегмента именно богатство и точность атрибутов определяют, насколько эффективно работают даже самые сложные модели.

Качественный поиск требует тщательной работы с

данными, что напрямую связано с инженерными сложностями трансформации. В DIY-сегменте важны признаки, связанные с техническими характеристиками (размеры, материалы, стандарты совместимости) [11]. Feature engineering включает текстовые, поведенческие, бизнес-признаки и session features, которые моделируют последовательности действий пользователей [12].

Сегмент DIY характеризуется наличием значительно более обширного набора технических атрибутов и спецификаций по сравнению с категориями fashion или grocery. Эта особенность предоставляет уникальную возможность для обогащения векторных представлений (эмбеддингов) и их непосредственной интеграции в модели машинного обучения (ML) в качестве высокоинформативных признаков [11]. Например, Ozon использует технические атрибуты для улучшения релевантности DIY-запросов, таких как «шурпы для гипсокартона», что повысило конверсию на 10% [17]. Учет session features улучшает релевантность для многошаговых задач, таких как поиск дрели, затем свёрл и крепежа. Эти аспекты подготовки данных формируют основу для экспериментального протокола.

VII. ЭКСПЕРИМЕНТАЛЬНЫЙ ПРОТОКОЛ

Для обеспечения объективности проверок критически важно сочетать оффлайн-оценку с онлайн-экспериментами. Такой подход позволяет минимизировать риски внедрения и одновременно ускорить цикл улучшений.

Для проверки гипотез применяются оффлайн- и онлайн-методы оценки. Оффлайн-оценка использует метрики $nDCG@K$ и $expected_revenue@K$ [3]. Для коррекции смещений в implicit feedback применяются методы IPS и Doubly Robust [12]. A/B-тесты измеряют влияние изменений на CTR, conversion rate и выручку [9].

Длительное проведение A/B-экспериментов (2–8 недель) необходимо для учета сезонности и вариаций поведения пользователей в DIY-сегменте [6]. План экспериментов включает определение метрик, выбор групп, настройку тестового окружения, мониторинг в реальном времени и статистический анализ результатов [9]. Абляционные исследования изолируют влияние компонентов, таких как мультимодальные эмбеддинги [11]. Этот протокол связан с наблюдениями и гипотезами, которые определяют направления дальнейших исследований.

Таким образом, разработанный протокол позволяет эмпирически проверять гипотезы и сопоставлять результаты различных архитектурных решений. На основе этих экспериментов формируются наблюдения и исследовательские гипотезы, представленные далее.

VIII. НАБЛЮДЕНИЯ И ИССЛЕДОВАТЕЛЬСКИЕ ГИПОТЕЗЫ

Приведенные примеры демонстрируют, что гибридные архитектуры работают в разных бизнес-контекстах и масштабах, однако детали реализации, cost model и скорость итерации отличаются.

- Amazon. Последовательные улучшения рекомендательных систем и переход к гибридным подходам позволили платформе масштабировать персонализацию и справляться с long-tail; в практических кейсах отмечено уменьшение latency и рост качества выдачи при внедрении векторных индексов [5, 6].
- Walmart. Semantic ads retrieval с LLM-предобучением показал значительное улучшение nDCG (примерно +16% в опубликованном эксперименте) для рекламных объявлений, что указывает на потенциал LLM/semantics в коммерческих задачах поиска [27].
- Wayfair. Публикация WANDS стимулировала исследования в области оценки релевантности и дала датасет для академических и промышленных сравнений моделей [45].
- Home Depot. Инвестиции в NER и structured extraction улучшили обработку технических запросов и, как следствие, повысили конверсию в категориях DIY [19].
- Alibaba / Taobao. Комплексный подход (search + payments + trust) показал, что технология поиска тесно связана с бизнес-экосистемой и может менять рыночную динамику [24, 25].
- Ozon / Wildberries. Локальные игроки адаптировали мировые практики к языковым и логистическим особенностям рынка, применяя гибридные подходы и cloud-scaling для обработки пиковых нагрузок [16, 17].

Данное исследование фокусируется на DIY-сегменте и task-oriented intent, предлагая моделировать последовательности действий пользователей через session sequences и использовать мультимодальные эмбеддинги [11, 12]. В результате формулируются следующие гипотезы:

- H1: Session-label модели превосходят pointwise LtR в обработке последовательных запросов в DIY [1].
- H2: Мультимодальные эмбеддинги повышают recall для технических запросов [4].
- H3: Revenue-aware ранжирование увеличивает LTV без потери пользовательского опыта [3].
- H4: Поэтапная трансформация снижает риски и улучшает метрики в DIY [10]. Данная гипотеза является ключевой для brown field проектов, поскольку постепенное внедрение гибридных архитектур и A/B-тестирование позволяют минимизировать downtime и

потерю функциональности, тем самым снижая бизнес-риски, связанные с миграцией от legacy-систем.

Эти гипотезы основаны на вызовах e-commerce поиска, где традиционные методы уступают гибридным. DIY-запросы, такие как «шурпы для гипсокартона», требуют учета технических атрибутов, что усложняет поиск по сравнению с fashion [28] или FMCG [29].

На основе этих наблюдений можно сделать вывод, что гипотезы H1–H4 не только отражают текущие тенденции, но и предлагают эмпирически проверяемые направления для дальнейших исследований, где интеграция мультимодальных эмбедингов с revenue-aware оптимизацией может стать ключом к повышению LTV в DIY-сегменте [3, 11]. Такой подход обеспечивает баланс между теоретическим анализом и практическими рекомендациями, подчеркивая эволюцию от реактивных систем к проактивным, адаптирующимся к пользовательским сессиям [12].

IX. ВЫВОДЫ И РЕКОМЕНДАЦИИ

В заключение можно отметить, что развитие поиска в DIY-e-commerce требует синтеза классических методов и современных ML-подходов. Новизна исследования подтверждается: 1) разработкой поэтапной стратегии трансформации, адаптированной к специфике DIY, и 2) предложением экспериментального протокола, связывающего теоретическую модель с эмпирической проверкой. Сравнение различных стратегий показывает, что гибридные архитектуры обеспечивают наилучший баланс между скоростью и качеством. В будущем особое значение приобретут мультимодальные методы и session-aware модели, способные адаптироваться к потребностям пользователя в реальном времени.

Для сегмента DIY наилучшей практикой является гибридная архитектура поиска, сочетающая оффлайн pre-ranking для популярных запросов, двухэтапное онлайн-ранжирование и векторный поиск для long-tail запросов. Перспективными направлениями развития являются session-aware модели, мультимодальные эмбединги и revenue-aware оптимизация. Реализация данных подходов позволит повысить качество поиска и улучшить бизнес-показатели интернет-магазинов.

- Модели с учетом сессионного контекста (Session-aware Models): Разработка и применение моделей, способных учитывать последовательность действий пользователя в рамках текущей сессии для более точного прогнозирования релевантности и интенций [12].
- Мультимодальное векторное представление (Multimodal Embeddings): Создание единых векторных пространств путем интеграции разнородных данных о товарах, включая текстовые описания, визуальный контент

(изображения) и технические спецификации [11].

- Оптимизация с ориентацией на доходность (Revenue-aware Optimization): Интеграция экономических метрик в целевую функцию ранжирования для оптимизации товарного портфеля по критерию маржинальности (валовой прибыли) [3].

Временные оценки перехода к LtR-ранжированию для DIY-ритейлера с каталогом до 1 млн SKU составляют 6–12 месяцев:

- Оценка и планирование (1–2 месяца): Аудит инфраструктуры и разработка roadmap [10].
- Подготовка данных (3–5 месяцев): Очистка, нормализация, разметка для LtR [13, 15].
- Внедрение гибридного поиска (2–3 месяца): Настройка keyword-based и векторного поиска [14].
- Оптимизация и тестирование (2–4 месяца): Тюнинг моделей и A/B-тесты [9].
- Масштабирование и мониторинг (1–2 месяца): Развертывание в production [10].

Важно отметить, что успешность трансформации определяется не только техническими решениями, но и организационными факторами: готовностью команды к экспериментам, наличием процессов непрерывной интеграции моделей и управлением изменениями в компании. Для DIY-сегмента значимым становится также обучение сотрудников работе с данными и формирование культуры data-driven решений. В перспективе можно ожидать усиления роли генеративных моделей, которые будут интегрироваться в поиск для интерактивного диалога с пользователем и решения комплексных задач.

Для крупных ритейлеров переход занимает 12–18 месяцев. Использование облачных сервисов может сократить сроки на 20–30% [13, 14]. Эти рекомендации обеспечивают конкурентное преимущество в DIY-сегменте.

В целом, проведенное исследование показывает, что переход к гибридным и векторным архитектурам в сегменте DIY является неотвратимым и необходимым условием для сохранения конкурентоспособности. Он требует баланса между точностью, скоростью и бизнес-целями, а также интеграции новых методов — от мультимодальных эмбедингов до session-aware моделей. Именно комплексное сочетание технологических и организационных мер позволит ритейлерам выстроить поисковые системы следующего

поколения.

Таким образом, развитие поисковых систем в DIY-сегменте можно рассматривать как постоянный поиск баланса: между точностью и скоростью, бизнес-интересами и пользовательским опытом, локальными решениями и масштабируемыми платформами.

БИБЛИОГРАФИЯ

- [1] Liu T.-Y. Learning to Rank for Information Retrieval Foundations and Trends in Information Retrieval, Vol. 3, No. 3, 2009, pp. 225–331. DOI: 10.1561/1500000016
- [2] Qin T., Liu T.-Y., Xu J., Li H. LETOR: A Benchmark Collection for Research on Learning to Rank for Information Retrieval. Information Retrieval Journal, 2010.
- [3] Wu L., Hu X., Shen D. Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce. Proceedings of ACM SIGIR, 2018. DOI: 10.1145/3209978.3209993
- [4] Johnson J., Douze M., Jégou H. Billion-scale similarity search with GPUs //IEEE Transactions on Big Data. – 2019. – Т. 7. – №. 3. – С. 535-547.
- [5] Linden G., Smith B., York J. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing, 2003.
- [6] Smith B., Linden G. Two decades of recommender systems at Amazon.com //Ieee internet computing. – 2017. – Т. 21. – №. 3. – С. 12-18.
- [7] Douze M. et al. The faiss library //arXiv preprint arXiv:2401.08281. – 2024.
- [8] The Home Depot. Home Depot introduces improved search for all. Corporate News, 2022. <https://corporate.homedepot.com/news/products/home-depot-introduces-improved-search-all> (дата обращения: 01.10.2025).
- [9] Wayfair Tech. Search & Recommender Systems. Wayfair Tech Blog, 2018–2022. <https://www.aboutwayfair.com/careers/tech-blog/wayfair-paper-accepted-at-recsys-2022-applies-state-of-the-art-method-to-generate-product-recommendations> (дата обращения: 01.10.2025).
- [10] Mozolevskiy D., AlShikh W. Comparative analysis of retrieval systems in the real world //arXiv preprint arXiv:2405.02048. – 2024.
- [11] Xu E. et al. Optimizing E-commerce Search: Toward a Generalizable and Rank-Consistent Pre-Ranking Model //Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2024. – С. 2875-2879.
- [12] Zhang X. et al. A Survey on Side Information-driven Session-based Recommendation: From a Data-centric Perspective //IEEE Transactions on Knowledge and Data Engineering. – 2025.
- [13] Hybrid Search for an E-Commerce Product Catalogue. Elastic Search Labs, 2024. <https://www.elastic.co/search-labs/blog/hybrid-search-e-commerce> (дата обращения: 01.10.2025).
- [14] Recipes to vectors: Building a hybrid search app with OpenSearch. OpenSearch Blog, 2025. <https://opensearch.org/blog/recipes-to-vectors-building-a-hybrid-search-app-with-opensearch/> (дата обращения: 01.10.2025).
- [15] What is vector search? Complete guide [2025]. Meilisearch Blog, 2025. <https://www.meilisearch.com/blog/what-is-vector-search> (дата обращения: 01.10.2025).
- [16] Ozon Tech Intro Meetup: как работают системы поиска, рекомендаций и рекламы Ozon Tech Blog, 2023. <https://habr.com/ru/companies/ozontech/articles/750196> (дата обращения: 01.10.2025).
- [17] Краснов Ф. В. Н1: гибридная система извлечения информации для поиска товаров в электронной торговле //Труды Института системного программирования РАН. – 2024. – Т. 36. – №. 5. – С. 227-240.
- [18] Cheng Y., et al. An End-to-End Solution for Named Entity Recognition in eCommerce Search. Proceedings of the AAAI Conference on Artificial Intelligence, 2021. URL (дата обращения: 01.10.2025).
- [19] Lin Z., et al. Research on the Evolution Path of Business Ecosystem of Platform Enterprises: A Case Study of Alibaba. Advances in Social Science, Education and Humanities Research, 2021. URL (дата обращения: 01.10.2025).
- [20] Kwong R., Lee E. W. J. Legitimacy building and e-commerce platform development in China: The experience of Alibaba. Technological Forecasting and Social Change, 2020. URL (дата обращения: 01.10.2025).
- [21] Yang L., et al. Personalized Embedding-based e-Commerce Recommendations at eBay. arXiv preprint, 2021. URL (дата обращения: 01.10.2025).
- [22] Ding, S., Attenberg, J., & Baeza-Yates, R. WANDS: Dataset for Product Search Relevance Assessment. Advances in Information Retrieval, Vol. 13185, 2022, pp. 128-141.
- [23] Wang Z. et al. Semantic Ads Retrieval at Walmart eCommerce with Language Models Progressively Trained on Multiple Knowledge Domains //arXiv preprint arXiv:2502.09089. – 2025.
- [24] Rahman M. D. E-commerce Evolution: A Strategic Analysis of Alibaba's Business Ecosystem //Available at SSRN 5109165. – 2024.
- [25] Се Юйин. О маркетинговой стратегии Alibaba / Се Юйин // Экономика и социум. – 2022. – № 6-1 (97). – С. 912-919. – EDN LZCMYY.
- [26] Yang, L., Cui, Y., Xuan, Y., et al. Personalized Embedding-based e-Commerce Recommendations at eBay. arXiv preprint arXiv:2102.06156, 2021. URL: <https://arxiv.org/abs/2102.06156>
- [27] Wang, Y., Guo, X., & Li, Y. Semantic Ads Retrieval at Walmart eCommerce with Language Models Progressively Trained on Multiple Knowledge Domains. arXiv preprint arXiv:2502.09089, 2025. URL: <https://arxiv.org/abs/2502.09089>
- [28] Ранжирование с персонализацией в fashion-ecom: архитектура, методы, эксперименты / А. С. Лысков, Н. Е. Лямин, А. И. Клещук, Е. А. Мамонова // Хабр. — 2024. — 21 августа. — URL: <https://habr.com/ru/companies/lamoda/articles/840370/> (дата обращения: 02.10.2025).
- [29] Luo C. et al. Exploring Query Understanding for Amazon Product Search //2024 IEEE International Conference on Big Data (BigData). – IEEE, 2024. – С. 2343-2348.
- [30] Google Shopping Case Studies. Think with Google. URL: <https://www.thinkwithgoogle.com/collections/shopping-case-studies/> (дата обращения: 02.10.2025).
- [31] These 11 Stores Are Killing It On Google Shopping (Case Studies). Store Growers. 2023. URL: <https://www.storegrowers.com/google-shopping-case-studies/> (дата обращения: 02.10.2025).

The Evolution of Search Engines in E-Commerce: From Static Catalogs to Hybrid and Vector Architectures

P.S. Astakhov, A.G. Belokrylov, F.V. Krasnov

Abstract—This article is devoted to the evolution of search systems in e-commerce with a particular focus on the DIY (Do-It-Yourself) segment. It examines the transition from static catalogs and SEO-based approaches to modern hybrid architectures that integrate Learning-to-Rank methods, vector search, and multimodal models. Special attention is given to engineering challenges related to scalability, data freshness, and the integration of machine learning into industrial solutions. The novelty of the research lies in the formulation of a step-by-step transformation strategy for search systems, specifically adapted to the requirements of the DIY segment and grounded in the analysis of practices adopted by leading retailers (Amazon, Walmart, Wayfair, Home Depot, Alibaba, Ozon, Wildberries). In addition, the study advances research hypotheses and proposes an experimental protocol that bridges theoretical modeling with empirical validation and objective evaluation of search quality. Thus, the article combines historical and applied analysis, offers a comprehensive perspective on the development of search technologies, and identifies promising directions for their application in the context of increasingly complex user queries and dynamic assortments.

Keywords — Hybrid search, Learning-to-Rank, Machine learning, Transformation, Information Retrieval, Search Engine Architecture

REFERENCES

- [1] Liu T.-Y. Learning to Rank for Information Retrieval. Foundations and Trends in Information Retrieval, Vol. 3, No. 3, 2009, pp. 225–331. DOI: 10.1561/1500000016.
- [2] Qin T., Liu T.-Y., Xu J., Li H. LETOR: A Benchmark Collection for Research on Learning to Rank for Information Retrieval. Information Retrieval Journal, 2010.
- [3] Wu L., Hu X., Shen D. Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce. Proceedings of ACM SIGIR, 2018. DOI: 10.1145/3209978.3209993.
- [4] Johnson J., Douze M., Jégou H. Billion-scale similarity search with GPUs //IEEE Transactions on Big Data. – 2019. – T. 7. – №. 3. – C. 535–547.
- [5] Linden G., Smith B., York J. Amazon.com recommendation: item-to-item collaborative filtering. IEEE Internet Computing, 2003.
- [6] Smith B., Linden G. Two decades of recommenders systems at Amazon.com //Ieee internet computing. – 2017. – T. 21. – №. 3. – C. 12–18.
- [7] Douze M. et al. The faiss library //arXiv preprint arXiv:2401.08281. – 2024.
- [8] The Home Depot. Home Depot introduces improved search for all. Corporate News, 2022. <https://corporate.homedepot.com/news/products/home-depot-introduces-improved-search-all> (date of access: 01.10.2025).
- [9] Wayfair Tech. Search & Recommender Systems. Wayfair Tech Blog, 2018–2022. <https://www.aboutwayfair.com/careers/tech-blog/wayfair-paper-accepted-at-recsys-2022-applies-state-of-the-art-method-to-generate-product-recommendations> (date of access: 01.10.2025).
- [10] Mozolevskiy D., AlShikh W. Comparative analysis of retrieval systems in the real world //arXiv preprint arXiv:2405.02048. – 2024.
- [11] Xu E. et al. Optimizing E-commerce Search: Toward a Generalizable and Rank-Consistent Pre-Ranking Model //Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2024. – C. 2875–2879.
- [12] Zhang X. et al. A Survey on Side Information-driven Session-based Recommendation: From a Data-centric Perspective //IEEE Transactions on Knowledge and Data Engineering. – 2025.
- [13] Hybrid Search for an E-Commerce Product Catalogue. Elastic Search Labs, 2024. <https://www.elastic.co/search-labs/blog/hybrid-search-e-commerce> (date of access: 01.10.2025).
- [14] Recipes to vectors: Building a hybrid search app with OpenSearch. OpenSearch Blog, 2025. <https://opensearch.org/blog/recipes-to-vectors-building-a-hybrid-search-app-with-opensearch/> (date of access: 01.10.2025).
- [15] What is vector search? Complete guide [2025]. Meilisearch Blog, 2025. <https://www.meilisearch.com/blog/what-is-vector-search> (date of access: 01.10.2025).
- [16] Ozon Tech Intro Meetup: How Search, Recommendations, and Advertising Work Ozon Tech Blog, 2023. <https://habr.com/ru/companies/ozontech/articles/750196/> (date of access: 01.10.2025).
- [17] Krasnov F. V. H1: hybrid retrieval system for product search in e-commerce //Proceedings of the Institute for System Programming of the RAS. – 2024. – T. 36. – №. 5. – C. 227–240.
- [18] Cheng Y., et al. An End-to-End Solution for Named Entity Recognition in eCommerce Search. Proceedings of the AAAI Conference on Artificial Intelligence, 2021.
- [19] Lin Z., et al. Research on the Evolution Path of Business Ecosystem of Platform Enterprises: A Case Study of Alibaba. Advances in Social Science, Education and Humanities Research, 2021.
- [20] Kwong R., Lee E. W. J. Legitimacy building and e-commerce platform development in China: The experience of Alibaba. Technological Forecasting and Social Change, 2020.
- [21] Yang L., et al. Personalized Embedding-based e-Commerce Recommendations at eBay. arXiv preprint, 2021.
- [22] Ding, S., Attenberg, J., & Baeza-Yates, R. WANDS: Dataset for Product Search Relevance Assessment. Advances in Information Retrieval, Vol. 13185, 2022, pp. 128–141.
- [23] Wang Z. et al. Semantic Ads Retrieval at Walmart eCommerce with Language Models Progressively Trained on Multiple Knowledge Domains //arXiv preprint arXiv:2502.09089. – 2025.
- [24] Rahman M. D. E-commerce Evolution: A Strategic Analysis of Alibaba's Business Ecosystem //Available at SSRN 5109165. – 2024.
- [25] Xie Yuying. On Alibaba's Marketing Strategy // Xie Yuying // Economy and Society. – 2022. – No. 6-1 (97). – P. 912–919. – EDN LZCMMY.
- [26] Yang, L., Cui, Y., Xuan, Y., et al. Personalized Embedding-based e-Commerce Recommendations at eBay. arXiv preprint arXiv:2102.06156, 2021. URL: <https://arxiv.org/abs/2102.06156>
- [27] Wang, Y., Guo, X., & Li, Y. Semantic Ads Retrieval at Walmart eCommerce with Language Models Progressively Trained on Multiple Knowledge Domains. arXiv preprint arXiv:2502.09089, 2025. URL: <https://arxiv.org/abs/2502.09089>

- [28] Lyskov, A. S., Lyamin, N. E., Kleshchuk, A. I., & Mamono va, E. A. (2024, August 21). Ranking with personalization in fashion-e com: architecture, methods, experiments. Habr. Retrieved October 2, 2025, from <https://habr.com/ru/companies/lamoda/articles/840370/>
- [29] Luo C. et al. Exploring Query Understanding for Amazon Product Search //2024 IEEE International Conference on Big Data (BigData). – IEEE, 2024. – C. 2343-2348.
- [30] Google Shopping Case Studies. Think with Google. URL: <https://www.thinkwithgoogle.com/collections/shopping-case-studies/> (date of access: 02.10.2025).
- [31] These 11 Stores Are Killing It On Google Shopping (Case Studies). Store Growers. 2023. URL: <https://www.storegrowers.com/google-shopping-case-studies/> (date of access: 02.10.2025).