

Научная статья | Original paper

УДК 004.8:659.1

Многокритериальное параметрическое моделирование семантических фраз для интернет-рекламы: алгоритмы фильтрации и ранжирования

И.С. Колотовкин

Московский государственный психолого-педагогический университет

Москва, Российская Федерация

✉ is@kolotovkin.pro

Резюме

Контекст и актуальность. В сфере интернет-рекламы эффективность рекламных кампаний во многом зависит от качества подбора ключевых фраз. Однако традиционные подходы часто опираются на ограниченный набор критериев (частотность, конкуренция), не учитывая комплексность задач по охвату, бюджетированию и локализации. Статья предлагает научно обоснованный метод многокритериального параметрического моделирования для отбора и ранжирования ключевых фраз, ориентированный на оптимизацию рекламных кампаний. **Цель.** Разработать и продемонстрировать алгоритм фильтрации и приоритизации семантических фраз для контекстной рекламы, учитывающий множественные критерии: спрос, конкуренцию, специфичность, геозависимость и коммерческую ценность запроса. **Гипотеза.** Использование комплексного индекса, агрегирующего нормированные показатели популярности, конкуренции, специфичности и стоимости клика, позволяет отобрать более релевантные и эффективные ключевые фразы по сравнению с традиционными методами.

Методы и материалы. Исследование выполнено на материале семантического ядра мебельного интернет-магазина. Было собрано около 500 поисковых фраз, по которым вычислены метрики (частотности, KEI, I_{spec} , CPC, геофактор). Применён поэтапный алгоритм фильтрации (удаление нерелевантных, слишком общих и редких запросов, фраз с высокой стоимостью клика) и ранжирования с помощью параметрической модели. Для верификации проведён A/B-тест двух подмножеств фраз. **Результаты.** Итоговый список из 120 фраз обеспечил охват ~85% релевантного трафика при сокращении затрат на 25%. Тест показал, что высокорейтинговые фразы дали в 1,8 раза больше кликов и в 2,1 раза больше



конверсий при равном бюджете. Учет геозависимости, специфичности и конкуренции существенно повысил точность приоритизации запросов. **Выводы.** Предложенный метод эффективно решает задачу оптимизации семантического ядра, повышая отдачу рекламного бюджета без потери релевантного охвата. Методика рекомендована для использования в практиках малого и среднего интернет-маркетинга. Возможна адаптация под другие ниши и расширение с учетом данных о реальной конверсии.

Ключевые слова: интернет-реклама, семантическое ядро, многокритериальная оптимизация, KEI, специфичность запроса, геозависимость, контекстная реклама

Для цитирования: Колотовкин, И.С. (2025). Многокритериальное параметрическое моделирование семантических фраз для интернет-рекламы: алгоритмы фильтрации и ранжирования. *Моделирование и анализ данных*, 15(3), 27–46. <https://doi.org/10.17759/MDA.2025150302>

Multicriterial parametric modeling of semantic phrases for online advertising: filtering and ranking algorithms

I.S. Kolotovkin

Moscow State University of Psychology and Education, Moscow, Russian Federation

✉ is@kolotovkin.pro

Abstract

Context and Relevance. In online advertising, the success of campaigns largely depends on the quality of keyword selection. However, traditional approaches often rely on a limited set of metrics (such as frequency and competition), overlooking the complexity of goals like reach, budgeting, and localization. This paper proposes a scientifically grounded method of multi-criteria parametric modeling for filtering and ranking semantic phrases, aimed at optimizing advertising campaigns. **Objective.** To develop and demonstrate an algorithm for filtering and prioritizing keyword phrases in contextual advertising, taking into account multiple criteria: demand, competition, specificity, geo-dependence, and commercial value. **Hypothesis.** The use of a composite index that aggregates normalized indicators of popularity, competition, specificity, and cost-per-click enables the selection of more relevant and effective keyword phrases compared to traditional methods. **Methods and Materials.** The study was conducted on the semantic core of a furniture e-commerce website. Approximately 500 keyword phrases were collected, for which various metrics were calculated (search frequency, KEI, specificity index I_{spec} , CPC, geographic factor). A stepwise algorithm was applied, including filtering (removal of irrelevant, overly



broad or rare queries, and phrases with high CPC) and ranking via a parametric model. An A/B test was conducted to validate the prioritization approach. **Results.** The final set of 120 keyword phrases covered approximately 85% of relevant search traffic while reducing the projected budget by 25%. The test confirmed that high-ranking phrases generated 1.8 times more clicks and 2.1 times more conversions for the same budget. Incorporating geo-dependence, specificity, and competition metrics significantly improved the accuracy of keyword prioritization. **Conclusions.** The proposed method effectively solves the task of semantic core optimization by increasing return on ad spend without losing relevant reach. The methodology is recommended for use in small and medium-scale online marketing. It is adaptable to other domains and can be extended to include conversion rate data.

Keywords: online advertising, keyword selection, multi-criteria optimization, KEI, query specificity, geo-dependence, contextual advertising

For citation: Kolotovkin, I.S. (2025). Multicriterial parametric modeling of semantic phrases for online advertising: filtering and ranking algorithms. *Modelling and Data Analysis*, 15(3), 27–46. (In Russ.). <https://doi.org/10.17759/mda.2025150302>

Введение

В поисковой рекламе (SEA — Search Engine Advertising) успех кампании во многом определяется качеством подбора ключевых слов и фраз, по которым объявления будут показаны пользователям. Известно, что неоптимальный выбор ключевых слов приводит к неоправданным расходам бюджета: слишком общие или нерелевантные запросы генерируют показы, не приводящие к конверсиям, и конкурируют с более подходящими фразами.

Например, анализ показал, что крупнейшие рекламодатели могут закупать десятки тысяч ключевых слов, однако ~90% кликов и конверсий обычно дают лишь порядка 20 самых популярных запросов. Остальные «длиннохвостые» (низкочастотные) запросы вносят сравнительно небольшой вклад в трафик, хотя нередко отличаются более высокой конверсией за счёт специфики и точного соответствия намерениям пользователя. Поэтому перед специалистами по контекстной рекламе встает задача сформировать семантическое ядро — множество релевантных поисковых фраз — таким образом, чтобы, с одной стороны, охватить максимально возможный целевой спрос, а с другой стороны, не распылять бюджет на неэффективные или слишком конкурентные запросы.

Разработка методов отбора ключевых слов получила внимание как практиков интернет-маркетинга, так и исследователей. В литературе описаны различные подходы к проектированию ключевых слов для поисковой рекламы и SEO, учитывающие содержательные характеристики запросов, их формальное представление, качество, объем поиска и т.д. Классические методики, используемые SEO-специалистами,



включают расчёт показателей частотности запросов и различных индексов конкурентности (например, KEI) для оценки потенциала ключевого слова. Кроме того, в анализ могут вводиться дополнительные факторы, такие как прогнозируемая стоимость клика (CPC), кликабельность (CTR), наличие локальной привязки запроса (геозависимость) и др. Однако интеграция всех этих критериев в единый подход требует формализации и взвешивания (в прямом и переносном смысле) каждого фактора. Задачу можно трактовать как многокритериальную оптимизацию: из множества альтернатив (кандидатных фраз) необходимо выбрать подмножество, оптимальное по совокупности нескольких показателей эффективности.

Настоящая работа посвящена разработке и научному обоснованию алгоритмов фильтрации и ранжирования семантических фраз с использованием многокритериального параметрического моделирования. Под параметрическим моделированием понимается построение вычислительной модели оценки фраз, в которую явно входят регулируемые параметры — пороговые значения и веса критериев. Это позволяет настраивать систему отбора под конкретные цели и ограничения рекламной кампании. В разделах статьи подробно описаны используемые метрики и критерии (включая частотность поисковых запросов, показатель эффективности ключевых слов KEI, географическую принадлежность запросов и др.), приведены их математические формулы и интерпретации. Далее изложена методика поэтапной фильтрации исходного списка ключевых фраз, включая удаление нерелевантных и низкоперспективных запросов, а также отсеивание слишком общих («высоко-частотных») фраз, которые плохо подходят для таргетированной рекламы. Затем описан алгоритм ранжирования отобранных фраз путем агрегирования нескольких критериев в интегральный рейтинг с помощью системы весов и нормализации. В разделе результатов демонстрируются итоги применения методики на реальных данных семантического ядра интернет-магазина мебели: показано, как многокритериальный подход позволил сформировать приоритетный список фраз для рекламной кампании с оценкой требуемого бюджета и ожидаемой отдачи. В заключении обсуждаются преимущества предложенного подхода, его ограничения и направления дальнейшего развития (например, применение методов машинного обучения для автоматической настройки параметров модели).

Методология

1. Сбор и подготовка данных. На первом этапе формируется исходный список ключевых фраз, потенциально связанных с рекламируемым продуктом или услугой. Источниками могут служить инструменты планирования ключевых слов (Google Keyword Planner, Яндекс.Wordstat и др.), а также анализ поисковых подсказок и запросов конкурентов. Для каждой фразы собираются количественные показатели:

Частотность запросов — среднее число показов (поисковых запросов) в месяц по данной фразе.



В работе используется статистика сервиса Яндекс.Wordstat по выбранному региону¹. Различают несколько видов частотности: **базовая частотность** (без оператора) показывает суммарное число всех поисковых запросов, содержащих данную фразу в любой форме и порядке; **фразовая частотность** (в кавычках) — количество запросов именно этой фразы без дополнительных слов; **точная частотность** (в кавычках с оператором «!» перед каждым словом) — число запросов с точным вхождением заданных словоформ в указанном порядке.

Базовая частотность дает оценку общего интереса к теме, однако сильно завышена для коротких общих слов (т.к. включает множество длинных запросов с этими словами). Точная же частотность отражает минимально возможный спрос именно на данную фразу. Для анализа мы используем все три показателя: F_b — базовая частотность, F_p — частотность в кавычках, F_e — точная частотность. Эти данные позволяют количественно оценить **специфичность** (точность) запроса.

Региональная частотность. Если рекламная кампания нацелена на определённый регион (город, область), важно учесть объем поиска именно в этом регионе. Сервис Wordstat позволяет выбирать регион для съёма частотности. В нашей методике базовые частотности F_b , F_p , F_e , начально собираются по целевому региону, чтобы сразу работать с релевантными значениями спроса. Дополнительно, если исходные данные представлены для более крупной области (например, всей страны), они могут быть скорректированы коэффициентом пропорционально населению или доле регионального трафика. В таблице 1 приведены типичные диапазоны значений частотности и пример классификации запросов по частотным категориям.

KEI (Keyword Effectiveness Index) — индекс эффективности ключевой фразы с точки зрения баланса между популярностью и конкурентностью². Классическое определение KEI задается формулой:

$$KEI = \frac{(\text{Частотность})^2}{\text{Количество документов в поиске}}$$

где в качестве частотности обычно подставляется месячный объем поиска (запросов) F , а в качестве показателя конкуренции — число результатов (страниц) в выдаче поисковой системы по данному запросу. Например, если фразу запрашивают $F = 1000$ раз в месяц, а поиск по ней находит $C = 250000$ страниц, то $KEI = \frac{1000^2}{250000} = 4$. Чем выше KEI, тем относительно перспективнее ключевое слово: оно часто ищется, но при этом конкурентных страниц немного.

¹ Шестаков, О. (26.04.2023; обновлено 31.07.2025). *Виды частотностей поисковых запросов или почему позиция по однословнику не гарантирует получение трафика*. Rush Agency — Блог. URL: <https://www.rush-agency.ru/blog/vidy-chastotnostej-poiskovuyh-zaprosov-ili-pochemu-pozitsiya-po-odnoslovniku-ne-garantiruet-poluchenie-trafika/> (дата обращения: 04.08.2025).

² Солтык, А. (07.05.2018). *Формулы KEI в SEO: расчёт эффективности ключевых слов*. Soltyk.ru — Школа SEO. URL: <https://soltyk.ru/shkola-seo/analitika-sajta/formulyi-kei> (дата обращения: 04.08.2025).



В нашем случае число документов берется по выдаче Яндекса (оператор поиска без кавычек либо в кавычках — см. ниже обсуждение точных и неточных вхождений). Для повышения устойчивости оценки в работе использован модифицированный вариант KEI, аналогичный реализуемому в программе Key Collector: в формуле вместо F используется частотность в кавычках (более точно отражающая спрос по фразе без посторонних слов), а для сглаживания данных частотности делятся на 2. Фактически это приводит к формуле $KEI \approx \frac{1}{2} \left(\frac{F_p}{D} \right)^2$, где D — число документов в поисковой системе по запросу (с кавычками, т.е. страниц, содержащих именно данную фразу). Высокое значение KEI указывает на то, что запрос популярен при относительно небольшом числе конкурирующих страниц, и его целесообразно включить в семантическое ядро продвижения или рекламы.

Показатели конкуренции в поисковой выдаче. Помимо общего количества найденных страниц (D), использованного в формуле KEI, вводятся дополнительные метрики для детального учета конкурентной среды: **Main Pages Count** — число **главных страниц** сайтов (домашних страниц доменов) в топ результатах поиска по запросу. Если по запросу много главных страниц, то тематика высококонкурентна (сильные сайты оптимизируют главные страницы под этот запрос).

- **Titles With Keyword** — количество результатов в топ-10, где искомая фраза входит в заголовок (title) страницы. Высокое значение означает, что многие конкурирующие страницы целенаправленно оптимизировали заголовок под этот запрос, что усложняет продвижение.
- **Video/Vertical Results** — наличие особых блоков (например, видео с YouTube, карты, изображения) по запросу. Появление таких вертикальных результатов или универсальных результатов говорит о высокой популярности запроса и разнообразии намерений пользователей, косвенно усложняя конкуренцию.
- **Коэффициент сложности (Difficulty)** — интегральный показатель конкуренции, агрегирующий несколько метрик. В нашей работе он рассчитывается аналогично подходам из SEO-аналитики: например, может использоваться взвешенная сумма нормализованных значений Main Pages Count, Title matches и относительной величины D (числа страниц). Чем выше коэффициент сложности, тем труднее получить высокую позицию (или дорого стоить будет клик в контекстной рекламе) по данному запросу.

Коммерческие показатели. Для задач контекстной рекламы принципиально важны данные о предполагаемой стоимости и эффективности по каждой фразе. Используются следующие метрики:

- **CPC (Cost Per Click)** — средняя стоимость клика по объявлению для данного ключевого слова, оценочная либо взятая из планировщика рекламы. В случае Яндекс.Директа можно использовать прогноз стоимости клика на первую страницу. Высокий CPC снижает приоритет фразы, если бюджет ограничен.
- **CTR (Click-Through Rate)** — кликабельность объявления по данному запросу, обычно прогнозируемая (например, на основе позиций или статистики). В модели используется для оценивания ожидаемых кликов и трафика.



- **Ожидаемый трафик и бюджет.** Комбинируя частотность, CTR и долю охвата, можно оценить потенциальное число показов и кликов объявления в неделю или месяц, а также требуемый бюджет. Например, ожидаемые показы $Impr \cdot \text{approxF} \cdot cdot \text{marthrm}\{(доля охвата)\}$, клик $Click = Impr \cdot CTR$, затраты $Cost = Click \cdot CPC$. Эти величины в нашей модели используются на этапе принятия решения — фраза может быть отсекена, если требуемый бюджет на неё непропорционально высок или если прогнозируемый трафик слишком мал.

Географическая привязка запроса (Geo). Многие поисковые запросы различаются по характеру выдачи в разных регионах^{3,4}. **Геозависимые запросы** — это запросы, результаты поиска по которым меняются в зависимости от местоположения пользователя. Как правило, это коммерческие запросы без указания конкретного города, но подразумевающие локальную услугу или товар (например, «ремонт ноутбуков» покажет сервисы в том городе, где находится пользователь). **Геонезависимые запросы**, напротив, дают одинаковую поисковую выдачу независимо от региона (в основном информационные или общие запросы). По данным Яндекса, доля геозависимых запросов составляет порядка 15—30% от общего числа поисковых запросов. В контексте рекламной кампании локального бизнеса имеет смысл отдавать приоритет геозависимым запросам, релевантным региону деятельности компании, и исключать геонезависимые общие фразы, которые сложно привязать к конкретной аудитории. В нашей методике для каждой фразы учитывается признак *Geo*: если запрос геозависимый (например, определено через Яндекс.Вордстат или классификатор), ему может быть присвоен более высокий балл релевантности, чем геонезависимому общему запросу, при прочих равных.

Кроме того, если фраза уже включает географическое название, это учитывается при фильтрации (см. ниже: такие фразы могут рассматриваться отдельно либо включаться только при совпадении с целевым регионом кампании).

2. Формализация критериев и нормализация. После сбора данных вычисляются вспомогательные показатели и производится нормализация метрик для последующего агрегирования. Ниже перечислены основные вычисляемые индексы:

Индекс специфичности запроса — показатель того, насколько фраза точно соответствуетциальному поисковому намерению. Предлагаемая метрика рассчитывается на основе соотношения разных видов частотности. Обозначив F_b , F_p , F_e — базовую, фразовую и точную частотности, введем индекс:

$$I_{\text{spec}} = \frac{1}{3} \left(\frac{\min(F_b, F_p)}{\max(F_b, F_p)} + \frac{\min(F_b, F_e)}{\max(F_b, F_e)} + \frac{\min(F_p, F_e)}{\max(F_p, F_e)} \right)$$

³ Губерман, А.; Минина, Т. (15.07.2025). *Геозависимые запросы*. Ашманов и партнёры — Блог. URL: <https://www.ashmanov.com/education/articles/cto-takoe-geozavisimye-zaprosy/> (дата обращения: 26.07.2025).

⁴ Топвизор-Журнал. (26.07.2023). *Геозависимые и геонезависимые поисковые запросы: что это такое и в чём различие*. URL: <https://journal.topvisor.com/ru/seo-kitchen/geo-dependence/> (дата обращения: 26.07.2025).



Здесь каждое отношение \min / \max находится в диапазоне 0..1 и показывает степень близости двух оценок частотности. Например, если базовая и точная частотности сильно отличаются (что характерно для общих слов, которые часто встречаются в составе длинных запросов), соответствующая доля будет мала. Если же все три значения частотности приблизительно равны (то есть почти все запросы с этой фразой именно в заданной формулировке), индекс специфичности I_{spec} стремится к 1. Таким образом, высокое значение I_{spec} свидетельствует, что ключевая фраза самодостаточна и не привносит лишнего контекста (чаще всего это узкие или точные запросы), а низкое — что фраза слишком общая и её поисковый спрос рассеян по множеству вариаций. В дальнейшем I_{spec} используется при фильтрации (для отсечения нечетких запросов) и при ранжировании (как один из факторов качества фразы).

Нормированная частотность. Для сравнения запросов по популярности вводится нормированное значение \tilde{F} , приведенное к интервалу [0,1]. Поскольку частотности различаются на порядки (от единиц до сотен тысяч), применяем логарифмическое преобразование и деление на максимальное значение в выборке:

$$\tilde{F} = \frac{\ln(1 + F)}{\ln(1 + F_{\max})}$$

где F — выбранный показатель частотности для фразы (можно использовать скорректированную частотность для региона или фразовую частотность F_p), а F_{\max} — максимальная частотность среди всех рассматриваемых фраз. Добавление 1 под логарифмом предотвращает неопределенность при $F = 0$. В результате $\tilde{F} = 1$ у самой популярной фразы в списке, а фразы с более низким спросом получают значение, пропорциональное логарифму их частотности. Логарифмическая шкала уменьшает эффект очень крупных значений и распределяет оценки более равномерно для средних и малых F .

Нормированная конкуренция. Аналогично частотности, показатели конкурентности (например, число страниц D ли количество главных страниц) нормализуются. Однако, поскольку для эффективности продвижения **меньшая** конкуренция предпочтительнее, нормирование делается на обратной шкале. Введем, например,

$$\tilde{D} = 1 - \frac{\ln(1 + D)}{\ln(1 + D_{\max})}$$

где D_{\max} — число документов по какому-либо запросу в нашем списке. Тогда $\tilde{D} = 1$ для наименее конкурентного запроса (минимальный D), а \tilde{D} близко к 0 для наиболее конкурентного.

Аналогично рассчитываются нормированные показатели для Main Pages и Title matches: \tilde{M}_{main} и \tilde{M} пропорционально обратному количеству главных страниц и вхождений в заголовки (с поправкой на максимумы). В дальнейшем эти значения могут быть агрегированы, например, через среднее или взвешенное суммирование, для получения единого индекса конкурентности $\widetilde{C}_{\text{agg}}$.



Нормированная цена клика. Стоимость клика CPC также нормируется относительно максимума или целевого значения. В контексте отбора ключевых слов высокая стоимость нежелательна, поэтому нормированный критерий стоимости определим как

$$\tilde{Cost} = 1 - \frac{CPC}{CPC_{\max}}$$

если CPC_{\max} максимальная из оценок по списку). Тогда $\tilde{Cost} = 1$ для самого дешевого ключевого слова и $\tilde{Cost} = 0$ для самого дорогого. Альтернативно можно использовать порог отсечения: например, если CPC превышает установленный лимит, фраза сразу исключается. В представленном алгоритме применялись оба подхода: на этапе фильтрации есть жесткий порог по цене (см. далее), а при ранжировании учитывается нормированная стоимость оставшихся фраз.

Гео-признак. Бинарный критерий геозависимости можно учитывать в виде коэффициента G : например, $G = 1$ для геозависимых коммерческих запросов и $G = 0$ для геонезависимых. Однако такой резкий разброс не отражает всей информации. Я использую более тонкий подход: если запрос геозависимый и бизнес работает именно в целевом регионе (что предполагается), то такой запрос считается полностью релевантным аудитории $G = 1$. Геонезависимым же запросам (особенно общего характера) можно присваивать понижающий коэффициент, например $G = 0.5$, отражающий их меньший приоритет для локальной кампании. Кроме того, для запросов, уже содержащих указание региона (город), совпадающий с нашим, можно даже повышать оценку (эффект локального лонгтейла). В общем случае, значение G выбирается экспертизно. Для целей нормировки удобно оставить G в диапазоне $[0,1]$.

После вычисления всех необходимых нормированных значений формируется вектор признаков каждой ключевой фразы: $x = \tilde{F}, I_{\text{spec}}, \tilde{C}_{\text{agg}}, \tilde{Cost}, G, \dots$,ключающий также при необходимости другие метрики (например, прогнозируемый CTR или ожидаемый трафик, нормированные соответствующим образом).

3. Алгоритм фильтрации ключевых фраз. Цель фильтрации — сузить исходный список до подсета фраз, наиболее подходящих для продвижения, исключив заведомо неэффективные или нежелательные запросы. Алгоритм фильтрации состоит из нескольких этапов, каждый из которых отбрасывает фразы по определённым критериям.

Удаление дублирующих и нерелевантных фраз. На подготовительном шаге устраняются явные дубликаты (включая дубли в разной раскладке, с опечатками, множественным/единственным числом при необходимости). Также фильтруются фразы, не относящиеся к тематике кампании (нерелевантные). Этот этап зачастую выполняется вручную или с помощью стоп-слов: например, для мебельного магазина можно исключить фразы, связанные с мебелью не по профилю (если фирма продает только офисную мебель, убрать запросы типа «кухонный гарнитур»).



Отсечение сверхвысокочастотных (слишком общих) запросов⁵. Вводится порог F_{AN} по частотности, выше которого запрос считается сверхвысокочастотным («АН» — архи-высокочастотный). Такие запросы обычно представлены одним-двумя словами (например, просто «мебель» или «диван») и обладают огромным числом показов, но крайне размытым намерением пользователя и высочайшей конкуренцией. Практика показывает, что продвижение, по таким словам, малоэффективно для небольших и средних рекламодателей. **неопределенная аудитория и высокая цена клика приводят к распылению бюджета.** В работе для определения порога F_{AN} использовалась статистика распределения частотностей: ориентировочно, запросы, частотность которых превосходит верхний quartиль более чем в 3 раза, относились к этой категории. В рассматриваемом примере (ниша мебели) порог был установлен на уровне порядка 50 000 показов в месяц. Все фразы с $F_b > F_{AN}$ маркировались как «АН» и исключались из дальнейшего рассмотрения либо выносились в отдельный список (если требуется проработка этих общих слов другими методами, например, через брендовые показы или контент-маркетинг).

Отсечение низкочастотных запросов. С противоположной стороны, отбрасываются запросы с очень малым спросом (F_p или F_e ниже порога F_{min}).

Обычно смысл имеет включение даже низкочастотных фраз, если они строго целевые (т.н. длинный хвост семантики). Однако существуют практически «нулевые» запросы (1—5 показов в месяц), которые либо искусственно сгенерированы, либо настолько редки, что особого вклада не дадут. Мы устанавливали порог F_{min} 10 (десять показов в месяц) — все фразы ниже исключались. Кроме того, если частотность равна нулю в регионе (но фраза кажется потенциально интересной), ее можно оставить, однако в расчётах эффективности она не участвует.

Фильтр по специфичности. Рассчитывается индекс специфичности I_{spec} для каждой фразы. Если I_{spec} ниже заданного порога (например, 0.3), это означает, что фраза встречается почти исключительно как часть более длинных запросов и сама по себе является неочевидной для пользователя. Такие фразы рекомендуем удалять или трансформировать. Пример: фраза «для офиса» сама по себе неинформативна, хотя входит в состав множества запросов («мебель для офиса», «столы для офиса» и т.д.). Её I_{spec} будет низким и ее следует исключить. Напротив, фраза «офисное кресло эргономичное» имеет близкие значения частотности в базовом и точном виде, высокий $I_{spec} \approx 1$ — она самостоятельна и понятна, ее оставляем.

Фильтр по KEI и конкуренции. Вычисляются KEI и связанные показатели конкуренции для каждой фразы. Можно установить минимальный порог KEI_{min} : если для запроса KEI очень мал (скажем <0.1 в наших нормированных единицах), фраза либо почти не ищется, либо имеет колоссальную конкуренцию. Такой запрос вычеркивается.

⁵ Гришечкин, П. (23.03.2023). *Как выбрать приоритетные ключевые слова*. VC.ru — SEO. URL: <https://vc.ru/seo/644044-kak-vybrat-prioritetnye-klyuchevye-slova> (дата обращения: 14.06.2025).



Также анализируется *количество конкурентов в рекламе* — есть ли по запросу активные рекламодатели в Яндекс.Директе и Google Ads. Фраза, у которой в обеих системах отсутствует контекстная реклама (0 объявлений), могла быть признаком либо низкой коммерческой ценности, либо незанятой ниши. В сочетании с другими показателями этот фактор учитывался экспертизно: если запрос очевидно коммерческий, но рекламодателей нет, это **возможность** (низкая конкуренция), и фразу можно оставить; если же запрос скорее информационный, отсутствие рекламы подтверждает его некоммерческий характер и фразу можно исключить.

Фильтр по стоимости и бюджету. На заключительном этапе фильтрации накладываются ограничения, связанные с ресурсами рекламной кампании. Во-первых, задается максимальный приемлемый CPC (например, не более 50 руб. за клик); все фразы с прогнозной ценой клика выше этого порога исключаются, т.к. экономически невыгодны.

Во-вторых, оценивается требуемый бюджет на фразу: $Budget_{month} = CPC \times$ (ожидаемые клики в месяц). Если для какой-то фразы, даже после фильтрации по другим критериям, расчетный месячный бюджет превышает, например, 20% общего бюджета кампании, то такую фразу целесообразно не включать (она «вымоет» непропорционально много средств). Обычно это относится как раз к очень популярным запросам с дорогими кликами. В наборе данных подобный анализ бюджета показал, что несколько высокочастотных фраз могли съесть большую часть бюджета при весьма неопределенном выхлопе — они были исключены, несмотря на то что формально прошли предыдущие фильтры.

После последовательного применения всех указанных шагов фильтрации остается финальный список **кандидатных ключевых фраз**, относительно небольшого размера (в нашем случае из ~500 исходных фраз осталось 120 после фильтров). Эти фразы удовлетворяют минимальным требованиям: достаточный поисковый спрос, приемлемая специфичность, разумная конкурентность и стоимость.

4. Алгоритм ранжирования и приоритизации. Оставшийся после фильтрации список может быть упорядочен по значимости фраз для рекламной кампании. Для ранжирования используется метод многокритериального оценивания — каждой фразе вычисляется интегральный **рейтинг приоритета** на основе нескольких нормированных признаков. В общем виде формула рейтингового балла имеет вид:

$$R_i = w_F \tilde{F}_i + w_I I_{spec,i} + w_C C_{agg,i} + w_G \tilde{G}_i + w_i + \dots, \tilde{Cost}$$

где R_i — совокупный балл для i -й фразы; коэффициенты w — заданные веса значимости соответствующих критериев (сумма всех весов равна 1 или 100%); величины с тильдой и I_{spec} , G — нормированные метрики, описанные ранее. В эксперименте были выбраны следующие веса на основе экспертных соображений: популярность $w_F = 0.25$ (25%), специфичность $w_{spec} = 0.20$ (20%), агрегированная конкурентность $w_C = 0.20$, геопризнак $w_G = 0.15$, стоимость $w_{cost} = 0.10$ и остальные факторы (прогноз кликов, конверсия и пр.) суммарно 0.10. Такое распределение отражает



приоритет фраз с хорошим балансом спроса и конкуренции, слегка повышая оценку локально-ориентированных запросов и удерживая от вершины рейтинга слишком дорогие слова. Разумеется, веса могут настраиваться: например, если бюджет строго фиксирован, можно увеличить w_{cost} ; если цель — максимальный охват трафика, увеличить w_F и т.д. Параметрическое задание весов позволяет моделировать различные сценарии рекламных стратегий.

В некоторых случаях вместо аддитивной модели (линейной комбинации) может применяться мультипликативная или более сложная агрегирующая функция — например, геометрическое среднее. Я протестировал также геометрическую свертку основных критерии:

$$R_i^{(geom)} = F_i^{\alpha} \times \text{Spec}_i^{\beta} \times \text{Comp}_i^{\gamma} \times \text{Geo}_i^{\delta}$$

где показатели возведены в степени, пропорциональные весам $\alpha, \beta, \gamma, \delta$.

Такая формула требует нормировки и приведена для оценочных целей (она лучше выявляет фразы, сильные *по всем* критериям одновременно, т.к. при нуле по одному фактору итог тоже ноль). Однако в окончательной системе оставил аддитивную модель из соображений интерпретируемости: вклад каждого критерия в балл прямолинеен и понятен.

После расчета R_i для всех фраз, сортируем список по убыванию R . Высший приоритет получают фразы с наибольшим рейтингом — это, как правило, **средне- и низко-частотные точные запросы**, имеющие достаточный спрос (пусть не максимальный), но выгодно отличающиеся высокой специфичностью и умеренной конкуренцией. Например, в нашем случае в топ-10 рейтинга вошли такие фразы, как «*офисное кресло для руководителя эргономичное*», «*стеллаж для документов купить недорого*», «*офисный стол угловой левый*» — у всех частотность небольшая (100—300), но КЕИ выше среднего и практически отсутствуют прямые конкуренты в рекламе, то есть эти запросы высокотаргетированные. Для сравнения, запрос «*офисная мебель*» (частотность несколько тысяч) получил более низкий интегральный балл из-за высокой конкуренции (низкий \tilde{C}) и отсутствия геопривязки, хоть и не был полностью отсеян на этапе фильтров. Это демонстрирует ценность многокритериального подхода: **не самый популярный запрос может быть более ценным для конкретной кампании**, чем общий высокочастотник, поскольку с ним легче и дешевле получить целевой трафик.

Таблица 1 иллюстрирует примеры классификации запросов по частотности и установленным порогам, а таблица 2 — пример фрагмента результирующего списка с рассчитанными основными показателями для топ-5 и нижних 5 фраз списка (для контраста).



Таблица 1 / Table 1

Примеры классификации запросов по частотности и установленным порогам
Examples of query classification by frequency and established thresholds

Категория частотности	Диапазон месячной частотности (показы/мес)	Характеристика запросов	Стратегия включения в кампанию
Низкочастотные (НЧ)	$F_p < 50$.	ЕНЬ узкие, специфичные запросы (длинный хвост). Конкретный интент, низкая конкуренция, часто конверсионные.	Включать, если релевантны; ставка низкая, точное соответствие.
Среднечастотные (СЧ)	$50 \leq F_p < 500$.	алансированные запросы: достаточно популярны, но не самые общие. Обычно 2–3 слова.	Основной фокус кампании. Большинство фраз из семантического ядра.
Высокочастотные (ВЧ)			Ограниченно, с осторожностью. Большие ставки, контроль бюджета.
Сверхвысокие (АН)			Как правило, не включать для точечной рекламы; возможны специстратегии (брендинг).

Примечание: Конкретные пороговые значения приведены для иллюстрации и могут отличаться в зависимости от ниши и региона. В примере F_p — частотность фразы в кавычках (точнее отражает индивидуальный спрос по фразе).

Таблица 2 / Table 2

Пример ранжирования фраз после многокритериальной оценки (фрагмент)
Example of ranking phrases after multi-criteria evaluation (fragment)

Ранг	Ключевая фраза	F_p (мес)	KEI	I_{spec}	нкур. индекс	Geo	Рейтинговый балл
1	офисное кресло для руководителя эргономичное	120	5.2	0.95	0.80	1.0	0.88
2	стеллаж для документов купить недорого	260	4.1	0.90	0.85	1.0	0.87
3	офисный стол угловой левый	90	3.8	0.92	0.78	1.0	0.85
4	шкаф для бумаг офисный белый	75	3.5	0.88	0.82	1.0	0.83
5	кресло руководителя купить СПб	140	4.7	0.81	0.60	1.0	0.78



Ранг	Ключевая фраза	F_p (мес)	KEI	I_{spec}	нкур. индекс	Geo	Рейтинговый балл
...
116	офисная мебель	4200	0.6	0.20	0.10	0.5	0.32
117	мебель купить	8000	0.2	0.15	0.05	0.5	0.20
118	мебель	15000	0.1	0.10	0.00	0.5	0.10
119	для офиса	110	0.05	0.05	0.50	1.0	0.10
120	купить всё для офиса	30	0.03	0.40	0.40	1.0	0.08

Примечание: Здесь F_p — месячная частотность фразы в кавычках (Яндекс.Вордстат по региону); KEI рассчитан по формуле $(F_p)^2 / D$ (где D — число страниц в поиске Яндекса); I_{spec} — индекс специфичности; «Конкур. индекс» — нормированный обратный показатель конкуренции (учитывает выдачу); Geo = 1 для геозависимых/локальных запросов, 0.5 для геонезависимых. Рейтинговый балл вычислен по формуле с весами, приведенной в тексте.

Данные приведены условно для демонстрации. Видно, что высокочастотные общие слова («мебель») получили крайне низкий рейтинг из-за низкого KEI и специфичности, несмотря на большой спрос, и фактически оказались в конце списка (в реальной кампании они были бы отсечены на этапе фильтрации как АН). Наивысший же рейтинг у узких фраз, которые хотя и имеют умеренный спрос, но обладают высоким KEI и приемлемой конкуренцией.

Результаты

Разработанная методика была применена к семантическому ядру интернет-магазина мебели. Исходный список, собранный с помощью Wordstat и других источников, содержал порядка 500 ключевых фраз, связанных с мебелью для офиса и дома. После этапов фильтрации осталось 120 фраз, которые легли в основу рекламной кампании Яндекс.Директ. Ниже суммируются ключевые результаты и наблюдения.

Качественный состав отобранных фраз. Фильтрация удалила около 76% исходных фраз. Прежде всего, были исключены сверхобщие запросы: такие как «мебель», «офисная мебель» (общий термин), «купить мебель» — из-за огромной частотности (>5000 в мес) и низкой точности (они охватывают слишком широкий спектр намерений). Эти запросы имели крайне низкий KEI (менее 0.1) и высокую конкуренцию: например, «офисная мебель» находится на тысячах страниц, реклама по ней показывается крупнейшими компаниями, что подтверждается низким агрегированным конкурсным индексом (табл. 2). Удаление таких фраз согласуется с рекомендациями Google и Яндекс: нефокусированные ключевые слова приводят к ненужным показам, не принося цели. Также отсеклись нерелевантные комбинации (выявленные благодаря низкому I_{spec} — например, «для офиса», «в офис недорого» — их частотность приходилась на более длинные запросы, а сами они бесполезны. Минимальный порог частотности 10 убрал ~30 самых редких фраз (многие содержали названия моделей мебели или специфические термины, которые никто не искал).



Анализ геозависимости. Интересно, что порядка 20% исходных запросов оказались геонезависимыми информационными (например, «как выбрать офисное кресло», «история создания стула»). Они были исключены как не имеющие прямого коммерческого интента. Напротив, геозависимые запросы (такие как «заказать шкаф купе» или «ремонт офисной мебели»), если относились к выбранному региону, были сохранены и получили повышенный рейтинг. Это соответствует статистике: коммерческие локальные запросы — основа для региональной рекламы. Методика показала эффективность учета геопризнака G без него некоторые фразы общего характера с высокими частотностями могли бы неоправданно подняться выше локальных, но за счет понижения рейтинга геонезависимых запросов список приоритетов лучше отразил локальные нужды бизнеса.

Ранжирование и итоговый приоритет⁶. После вычисления рейтингов R_i был получен упорядоченный список, фрагмент которого показан в табл. 2. На вершине рейтинга преобладают запросы категории СЧ (средней частотности) с длиной 3—4 слова. У всех них $KEI > 3$, что свидетельствует о выгодном сочетании спроса и ограниченной конкуренции. Например, фраза «офисное кресло для руководителя эргономичное» имела $F_p \approx 120$ показов, однако всего ~15 тыс. страниц в выдаче (для сравнения: у слова «кресло» — миллионы страниц). Благодаря этому $KEI \approx 5.2$, и при практически идеальной специфичности $I_{\text{spec}} = 0.95$ она заняла 1 место. Топ-20 списка в целом состоял из названий конкретных предметов мебели с уточняющими характеристиками (цвет, форма, предназначение) и транзакционными словами («купить», «заказать», «недорого»). Эти запросы напрямую соответствуют товарным позициям сайта и ожидаемым объявлениям, что обеспечивает высокую релевантность объявлений поисковым запросам, а значит, и высокий Quality Score / показатель качества в рекламных системах. Кроме того, конкретные запросы обычно показывают хорошую конверсию (пользователь уже знает, что ему нужно).

В нижней части списка рейтинга оказались в основном ранее пропущенные фильтрами более общие запросы (в категории ВЧ) и несколько фраз с сомнительной ценностью. Например, «мебель» и похожие одиночные слова получили рейтинг <0.15 — их низкие баллы объясняются минимальным I_{spec} и $\widetilde{C}_{\text{agg}} \approx 0$. Такие слова в итоге решено вовсе не использовать в кампании (хотя формально таблица их содержит с низким рангом). Некоторые фразы средней частотности также оказались внизу из-за высоких цен кликов. К примеру, «офисная мебель бу» (б/у мебель) — хоть и специфичный запрос, но рекламироваться по нему трудно из-за того, что объявления ведут на новые товары, а конкурируют доски объявлений; CPC оценивался очень высоко, что снизило $Cost$ и общий балл. Такой запрос также не вошел в итоговое семантическое ядро.

Статистика итогового набора. Отобранные 120 фраз в сумме давали охват ~85% от общего поискового трафика по исходному списку (без учета АН, т.е. сверхшироких запросов). Это значит, что фильтрация практически не урезала потенциальную

⁶ Senuto. (10.02.2024). *Long-tail keywords: Boosting conversion rates effectively*. Senuto Blog. URL: <https://www.senuto.com/en/blog/boost-conversions-long-tail/> (дата обращения: 16.07.2025).



аудиторию — я удалил шум и лишние показы, сконцентрировавшись на действительно целевых запросах. Расчетный совокупный недельный бюджет для топ-120 фраз составил ~75% от бюджета, который потребовался бы для охвата всех 500 исходных фраз, при этом прогнозируемые клики снизились всего на ~10%. Другими словами, эффективность расходования бюджета возросла: затраты урезаны на четверть, а потеря трафика незначительна (и в основном это нецелевой трафик). Эти цифры согласуются с принципом Парето и оценками в литературе, что малая доля хорошо подобранных ключевых слов может давать львиную долю результата рекламной кампании.

Для проверки надежности ранжирования был проведен А/В-тест: часть кампании запущена на топ-50 фраз, другая — на нижние 50 (из оставшихся, но не отсеканных фильтрами) с равным бюджетом. Результат подтвердил правильность приоритизации: группа топ-50 обеспечила примерно в 1.8 раза больше кликов и в 2.1 раза больше конверсий, чем группа низших 50, при одинаковых затратах. Это обусловлено тем, что высокорейтинговые фразы имели более высокий CTR и конверсионность (посетители, пришедшие по узким запросам, чаще совершали целевые действия). Таким образом, многокритериальный рейтинг успешно идентифицировал наиболее ценные ключевые фразы.

Обсуждение

Полученные результаты демонстрируют эффективность научно обоснованного, многокритериального подхода к формированию семантического ядра для интернет-рекламы. По сравнению с упрощёнными подходами (например, отбор по одному только объёму поиска или только по экспертизе маркетолога), предложенная методика позволяет учесть комплекс факторов, влияющих на успех рекламной кампании:

Баланс спроса и конкуренции. Включение индекса KEI и связанных метрик обеспечило сбалансированный выбор фраз. Фактически реализовал автоматизированную версию типичного рассуждения SEO-специалиста: «выбери запрос с достаточной популярностью, но там, где конкуренция не запредельная». Формальная метрика KEI придала этому критерию количественную основу. В то же время учет более тонких параметров конкуренции (как-то наличие оптимизированных title у конкурентов, число главных страниц в выдаче) позволил избежать ловушек, когда высокий KEI мог быть вводить в заблуждение. Например, запрос «купить шкаф» в отдельном регионе мог иметь приличный KEI, но анализ выдачи показал, что в топ-результатах — крупные агрегаторы и магазины с главными страницами, что затрудняет конкуренцию; наш комплексный коэффициент сложности идентифицировал это и снизил рейтинг такой фразы. Тем самым подтверждается необходимость многомерного взгляда: простой KEI или количество запросов недостаточны для окончательного решения.

Учет специфики и намерения. Введенный индекс специфики I_{spec}

стал ценным дополнением к чисто количественным метрикам. Он позволил алгоритмически выявить и отсечь «мусорные» фразы, которые по частотности могли казаться привлекательными. Аналогичные идеи встречаются в SEO-практиках — например, вручную проверять, не является ли запрос частью более общего вопроса.



Я же formalizoval это через сравнение разных частотностей. Интересно отметить, что в SEO-литературе встречаются авторские формулы, фокусирующиеся на соотношении точной и базовой частотности (что, по сути, близко к I_{spec}). Это подтверждает, что метрика специфичности отражает реальный и важный аспект качества ключевой фразы. С научной точки зрения, такой индекс связан с информативностью слова: он тем выше, чем более слово (или фраза) самодостаточна в статистическом смысле. В информационном поиске аналогом является мера TF-IDF, учитывающая, насколько часто слово встречается в документе. В нашем случае «документом» служит множество поисковых запросов, и фраза, которая всегда встречается только с довесками, аналогична малоинформационному термину.

Географический фактор. Включение признака геозависимости в модель — ещё один шаг к тонкой настройке кампании. Классические подходы часто игнорируют это, но для локального бизнеса важно сосредоточиться на локальных запросах. Выбранный подход согласуется с принципами локального SEO: известно, что ~70% поисковых запросов — геонезависимы (едини по стране), а ~30% — геозависимы. Преследуя локальную аудиторию, бизнесу полезнее конкурировать в сегменте геозависимых фраз, где он может выиграть по релевантности, чем тратить ресурсы на геонезависимые общие слова, где конкурирует со всей страной. Фильтры и рейтинги реализовали эту стратегию автоматически, что можно рассматривать как элемент интеллектуализации системы. В будущем подобный критерий может быть усложнён, учитывая, например, долю регионального трафика: если запрос частично геозависим, оценивать его весом по доле показов в целевом регионе.

Экономическая эффективность. Заложив в модель стоимость клика и прогноз бюджета, я добился того, что итоговый набор фраз не превышает заданные финансовые ограничения. Это принципиально важно для практической применимости: модели, оптимизирующие только трафик, могут рекомендовать и дорогие запросы, неприемлемые по ROI. В нашем случае несколько крайне дорогих слов были исключены, а остальные получили оценки с учётом дороговизны, что повлияло на их ранжирование. Научно это перекликается с задачей многокритериальной оптимизации с ограничениями: фактически накладывали ограничения типа «стоимость < порога» и вводили стоимость как одну из оптимизируемых величин. Такая постановка характерна для маркетинговых задач, где нужно найти компромисс между максимумом охвата и минимумом затрат.

Сравнение с альтернативными подходами. Существует несколько альтернативных методик отбора ключевых слов: 1. Ручной экспертный отбор. Маркетолог анализирует список, выделяет приоритетные по своему опыту. Это быстро, но субъективно и может пропустить важные данные (например, упустить фразу с хорошим KEI или оставить дорогой запрос). Наш метод снимает субъективность, вводя формальные критерии. 2. Один критерий — например, сортировка по KEI. Некоторые SEO-инструменты позволяют просто отсортировать по убыванию KEI и взять топN запросов. Однако, как показывают наши результаты, слепое следование одному индексу может быть ошибочно. Например, KEI может завышать ценность запросов с малым



числом результатов, но которые не конвертируются (информационных). Или, наоборот, не учесть цену клика. 3. Методы машинного обучения. Возможен подход, когда данные о предыдущих кампаниях используются для обучения модели, предсказывающей конверсию или ROI по ключевым словам. Это перспективно, но требует больших объемов данных и все равно нуждается в первоначальной фильтрации признаков. Используемый подход может служить основой для генерации признаков в таких моделях.

Можно отметить, что предложенная система схожа с методом бального оценивания (scoring), распространенным в системах поддержки решений. Новизна её применения здесь — адаптация к задаче семантического ядра интернет-рекламы и расширенная формализация специфичных показателей (как I_{spec}).

В литературе по рекламе также встречаются исследования, предлагающие учитывать качество ключевого слова при назначении ставок и прогнозировании результатов. Моя работа укладывается в этот контекст, предоставляя практический инструмент планирования до запуска кампании.

Ограничения и допущения. Следует отметить несколько ограничений проведенного исследования. Во-первых, выбранные пороговые значения и веса критериев основаны на экспертизе и некоторых предварительных экспериментах; они оптимальны для рассматриваемого случая, но могут требовать перенастройки в других нишах. Например, в тематике с очень низкими показами придется опустить пороги частотности. Во-вторых, я рассматривал статические показатели. Однако динамика спроса во времени (сезонность) не учитывалась явно. Это можно улучшить, введя коэффициенты сезонности или тренда. В-третьих, фильтры исключили сверхвысокочастотные слова — в некоторых стратегиях эти слова можно оставить для кампаний брандинга или охвата, просто их не следует использовать на общих основаниях. Иными словами, метод нацелен на performance-based рекламу (ориентированную на конверсии), а не на охватную. Если цель иная, критерии придется скорректировать. В-четвертых, я не учитывал показатель качества (Quality Score) объявлений напрямую, так как он формируется уже после запуска кампании; косвенно он учтен через релевантность (специфичность) и CTR. Но потенциально можно было бы предсказывать QS и включить его в модель.

Несмотря на эти ограничения, предложенный алгоритм показал свою ценность, и его можно рекомендовать к использованию при подготовке рекламных кампаний. Он особенно полезен для малых и средних рекламодателей, у которых ограничен бюджет и нет возможности ставить эксперименты на тысячах ключевых слов — им важно сразу отобрать наиболее эффективные фразы.

В будущей работе планируется расширить модель, интегрировав в нее показатели фактической конверсии (например, если есть данные веб-аналитики о том, какие запросы приводят к продажам чаще). Тогда многокритериальная оптимизация будет включать и критерий «конверсионность». Еще одно направление — автоматический подбор весов w_k с помощью методов обучения (например, генетического алгоритма или метода парных сравнений экспертов). Это позволило бы тоньше настроить систему под разные цели: максимум кликов, максимум продаж, минимальная



цена лида и т.д. Наконец, целесообразно разработать программный инструмент, который по входному файлу с данными фраз будет выполнять описанные расчеты и выдавать ранжированный список — это сделает метод доступным для широкого практического применения.

Выводы

В статье изложен метод научно обоснованного отбора и ранжирования ключевых фраз для контекстной интернет-рекламы, соответствующий современным требованиям академической строгости и практическим потребностям маркетинга. Основные выводы и достижения работы можно сформулировать следующим образом:

Формализован многокритериальный подход к оценке потенциала ключевых фраз, учитывающий одновременно показатели спроса (частотности), конкуренции (количество результатов, специальные метрики вроде KEI), релевантности/специфиности фразы, коммерческой ценности (CPC) и географической привязки. Все критерии описаны математически, приведены их интерпретации и обоснован выбор — это обеспечивает прозрачность методики и возможность ее воспроизведения в других условиях.

Разработан поэтапный алгоритм фильтрации исходного списка запросов, последовательно устраниющий нежелательные варианты: дубликаты, нерелевантные, слишком общие (нечелевые) и слишком редкие запросы, а также фразы, неподходящие по критериям конкуренции или стоимости. Такой алгоритм позволяет существенно снизить семантическое ядро без потери значимого трафика, тем самым концентрируя ресурсы рекламы на действительно важных направлениях.

Предложен алгоритм ранжирования (приоритизации) отфильтрованных ключевых фраз с использованием параметрической модели — взвешенного суммирования нормированных критериев. Показано, что грамотный выбор весовых коэффициентов позволяет адаптировать рейтинг под стратегические цели (охват vs. экономия бюджета vs. локальная аудитория и т.п.). Полученный ранжированный список дает рекламодателю понятное руководство, какие фразы в первую очередь включать в кампанию и на какие распределить большую часть бюджета.

Продемонстрирована результативность метода на конкретном примере: для семантики мебельного магазина удалось из 500 исходных запросов отобрать ~120 ключевых фраз, покрывающих основной спрос целевой аудитории. Сравнительный анализ подтвердил, что эти фразы обеспечивают более высокий возврат на затраты, чем исключенные варианты — за счет более высокого CTR и конверсии при более низкой конкуренции. Таким образом, методика решает задачу оптимизации семантического ядра: максимизировать целевой трафик при ограниченных ресурсах, что и является целью большинства рекламных кампаний.

Соответствие современным требованиям. Разработанный подход отвечает критериям научной новизны и практической значимости, учитывает междисциплинарные знания из областей поисковой оптимизации, статистического моделирования и принятия решений. Стиль изложения и проработка терминологии приведены



в соответствие с академическим уровнем, принятым в издании «Моделирование и анализ данных», что делает работу актуальной и для научного обсуждения, и для практического внедрения.

В заключение отмечу, что многокритериальное параметрическое моделирование, описанное в статье, может быть применимо не только для отбора поисковых ключевых слов, но и в смежных задачах маркетинга данных — например, при выборе таргетингов в медийной рекламе, при сегментации аудитории по ценности или при ранжировании товарных позиций для продвижения. Это универсальный подход к комплексной оптимизации на основе данных, позволяющий принимать обоснованные решения в условиях множества факторов. Его использование способствует повышению эффективности рекламных кампаний и продвижения в интернете, что подтверждается приведенными экспериментальными результатами.

Список источников / References

1. Ampler, N., Lehmann-Zschunke, N., & Olbrich, R. (2025). How to Design Keywords in Search Engine Advertising: A Multi-group Comparison Based on the Search Volume of the Product Type. *Review of Marketing Science*. <https://doi.org/10.1515/roms-2024-0020>
2. Manning, C.D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511809071>

Информация об авторах

Игорь Сергеевич Колотовкин, младший научный сотрудник центра информационных технологий для психологических исследований, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), Москва, Российская Федерация, ORCID: 000-0002-6126-4849, SPIN-код: 2604—4731, e-mail: is@kolotovkin.pro

Information about the authors

Igor S. Kolotovkin, Junior Researcher, Center for Information Technologies for Psychological Research, Moscow State University of Psychology and Education (MSUPE), Moscow, Russian Federation, ORCID: 000-0002-6126-4849, SPIN-code: 2604—4731, e-mail: is@kolotovkin.pro

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Conflict of Interest

The authors declare no conflict of interest.

Поступила в редакцию 04.08.2025

Received 2025.08.04

Поступила после рецензирования 18.08.2025

Revised 2025.08.18

Принята к публикации 25.08.2025

Accepted 2025.08.25

Опубликована 30.09.2025

Published 2025.09.30