

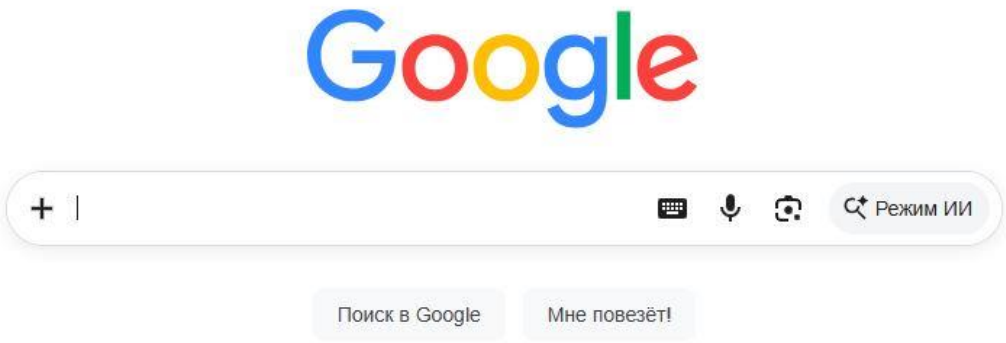

Изображение от jcompna Freepik

# СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ПОИСКА ИНФОРМАЦИИ

# СОДЕРЖАНИЕ

1. Поиск информации: как работают поисковые системы – 3
2. Формулирование поисковых запросов – 11
3. Базовые операторы поиска – 25
4. Продвинутые операторы и фильтры – 35
5. Логические операторы и комбинирование – 43
6. Поиск в специализированных областях – 55
7. Поиск в научных и академических базах данных – 65
8. Оценка достоверности источников – 75
9. Поиск в социальных сетях и на форумах – 90
10. Профессиональный поиск: право, медицина, техника – 102
11. Конфиденциальность и безопасность – 108
12. Создание персональной системы поиска - 120

# ПОИСК ИНФОРМАЦИИ:
# КАК РАБОТАЮТ ПОИСКОВЫЕ СИСТЕМЫ



# ПОИСК ИНФОРМАЦИИ: КАК РАБОТАЮТ ПОИСКОВЫЕ СИСТЕМЫ

## Поисковик не ищет

Когда вы вводите запрос и нажимаете Enter, поисковая система не отправляется в путешествие по интернету в поисках ответа. Он вообще не ищет в сети в реальном времени. Вместо этого он работает с огромной базой данных, которую построил заранее.

Эта база — результат трёх последовательных этапов: обхода, индексации и ранжирования.



# ПОИСК ИНФОРМАЦИИ: КАК РАБОТАЮТ ПОИСКОВЫЕ СИСТЕМЫ

## Три этапа: обход, индексация, ранжирование

Работа любой поисковой системы строится на трёх последовательных этапах. Обычно их называют обход (crawling), индексация (indexing) и выдача результатов (serving). Другие источники называют третий этап ранжированием (ranking).

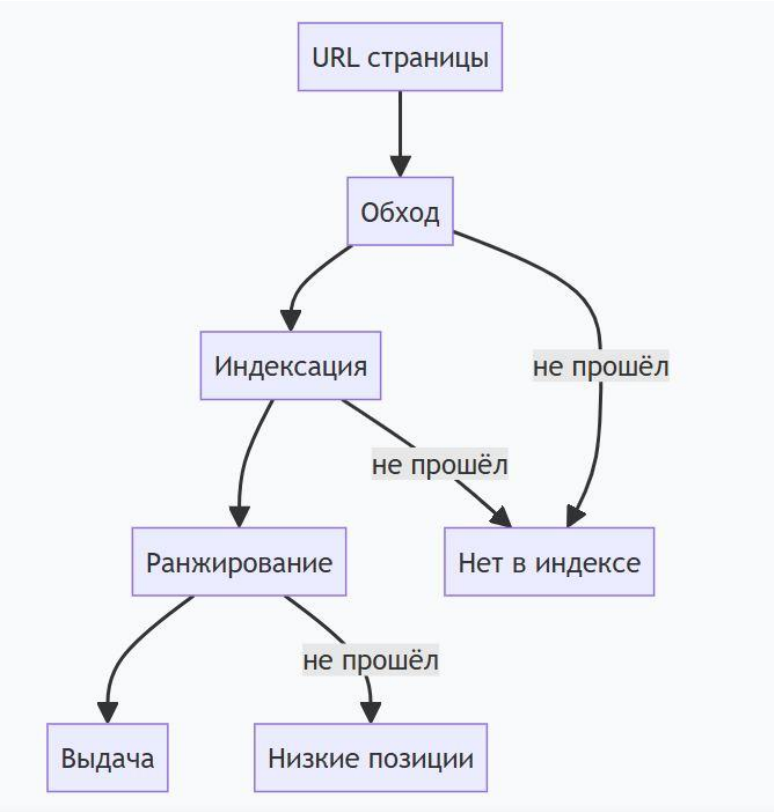
На этапе обхода автоматические программы-краулеры обнаруживают и скачивают страницы. Затем система анализирует содержимое, определяет каноническую версию и сохраняет информацию в огромной базе данных — индексе. И только когда пользователь вводит запрос, алгоритмы ранжирования выбирают из индекса наиболее релевантные страницы и упорядочивают их по полезности.

# ПОИСК ИНФОРМАЦИИ: КАК РАБОТАЮТ ПОИСКОВЫЕ СИСТЕМЫ

## Три этапа: обход, индексация, ранжирование

Здесь действует железный принцип: если страница не прошла предыдущий этап, следующий не наступает.

Нельзя ранжировать то, чего нет в индексе. Вы не сможете найти книгу, которую никогда не занесли в каталог, — сколько бы ни искали.



# ПОИСК ИНФОРМАЦИИ: КАК РАБОТАЮТ ПОИСКОВЫЕ СИСТЕМЫ

## Обход: как краулеры находят страницы

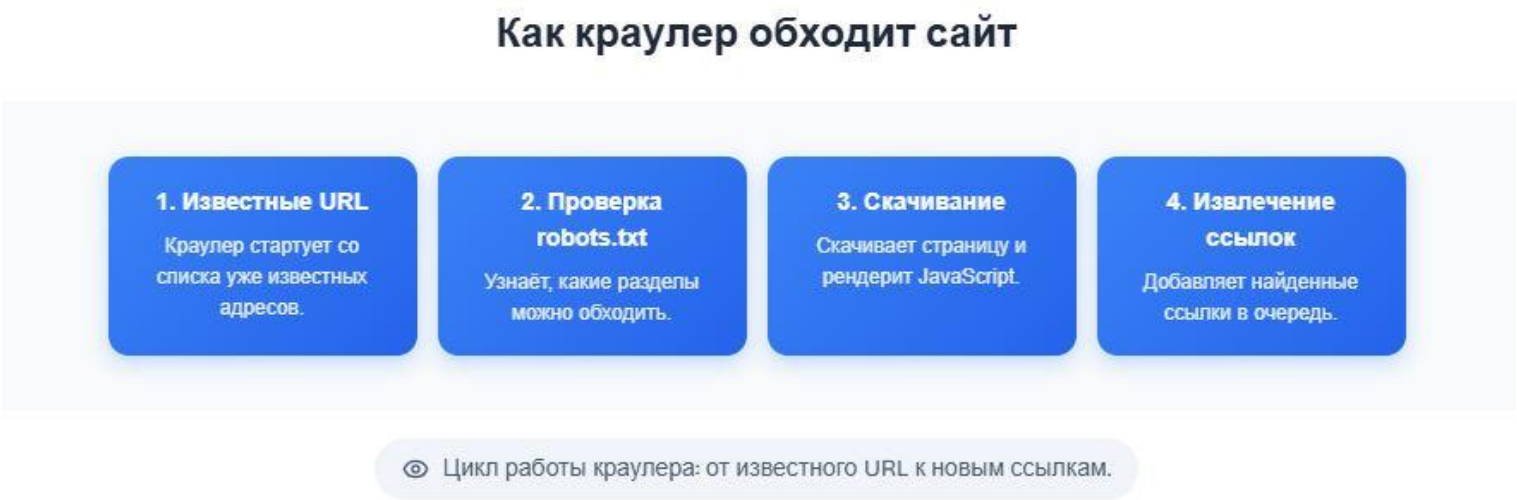
Обход — это первый этап, в ходе которого автоматические программы — краулеры, боты, пауки — систематически просматривают веб, скачивая содержимое страниц: текст, изображения, видео и ссылки. Google использует для этого программу Googlebot, которая работает на огромном множестве компьютеров и обрабатывает миллиарды страниц.

Краулеры начинают со списка известных URL, а затем переходят по ссылкам с этих страниц на новые, постоянно расширяя свой охват.

Обход не гарантирует, что страница будет проиндексирована. Это лишь доставка содержимого поисковику. Страницы, доступные только после входа в аккаунт, краулеры не обходят вовсе.

# ПОИСК ИНФОРМАЦИИ: КАК РАБОТАЮТ ПОИСКОВЫЕ СИСТЕМЫ

## Обход: как краулеры находят страницы




Цикл работы краулера: от известного URL к новым ссылкам.

# ПОИСК ИНФОРМАЦИИ: КАК РАБОТАЮТ ПОИСКОВЫЕ СИСТЕМЫ

## Индексация: как поисковик хранит данные

После того как страница скачана, поисковая система анализирует её содержимое и решает, стоит ли включать её в свою базу данных — индекс. Индексация включает обработку текста, ключевых тегов (например, title и alt), изображений и видео.

Система определяет, является ли страница дубликатом другой, и если да, выбирает каноническую версию — ту, которая будет показываться в результатах поиска. Для этого похожие страницы группируются в кластеры, и из каждого кластера выбирается наиболее репрезентативная.

Индексация — это ещё и фильтр качества: страницы с дублирующимся контентом, запретами (noindex, robots.txt) или малым объёмом полезной информации могут быть исключены из индекса.

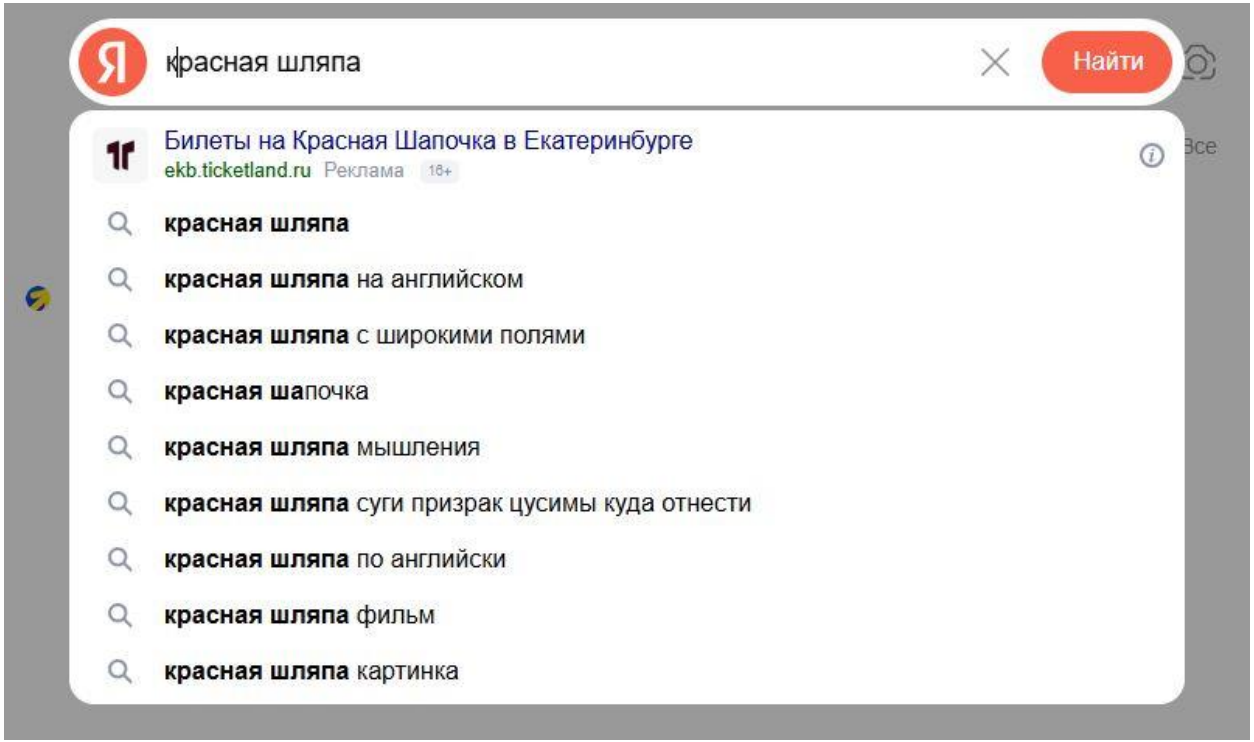
# ПОИСК ИНФОРМАЦИИ: КАК РАБОТАЮТ ПОИСКОВЫЕ СИСТЕМЫ

## Ранжирование: почему одни страницы в топе, а другие нет

Ранжирование — это третий этап, который запускается только после того, как пользователь вводит запрос. Поисковая система не просматривает весь интернет в реальном времени — она ищет в своём индексе страницы, соответствующие запросу, и сортирует их по степени полезности.

Алгоритмы ранжирования учитывают множество факторов: смысл запроса, релевантность контента, качество страницы, удобство для пользователя, авторитетность сайта, а также местоположение и настройки пользователя. Google заявляет, что использует более 200 факторов, но точный состав держится в секрете. Вес каждого фактора зависит от типа запроса: для новостных запросов важнее свежесть контента, чем для словарных определений.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ



# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Проблема словаря: почему поиск не находит

Центральная проблема, которую решает формулирование запросов, — это несоответствие между словарём пользователя и словарём, используемым в индексируемых документах. Один и тот же юридический или технический концепт может описываться десятками различных терминов в разных юрисдикциях и документах.

Разные методы поиска решают эту проблему по-разному. Булев поиск требует, чтобы вы заранее предугадали все возможные формулировки. Естественно-языковой поиск автоматически распознаёт концептуальные связи между разными терминами. Понимание этой проблемы — первый шаг к эффективному формулированию запросов, потому что именно она определяет выбор стратегии: точное соответствие терминов или смысловая интерпретация.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Проблема словаря: почему поиск не находит

### 🔍 Булев поиск

- ✅ Точное соответствие терминов
- ✅ Контроль над областью поиска
- ✅ Воспроизводимые результаты
- ❌ Нужно предугадать все формулировки
- ❌ Пропускает синонимы

### 💬 Естественно-языковой поиск

- ✅ Понимает смысл, а не слова
- ✅ Находит синонимы автоматически
- ✅ Прощает неточные формулировки
- ❌ Ранжирование непрозрачно
- ❌ Полноту оценить сложно

👁 Булев поиск даёт контроль, естественно-языковой — охват. Сильные стороны одного — слабости другого.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Булев поиск: контроль и воспроизводимость

Булев поиск (также называемый terms-and-connectors поиском) — это построение запроса из терминов и операторов. Операторы AND, OR, NOT объединяют концепты, операторы близости требуют нахождения слов в пределах заданного расстояния, кавычки фиксируют точную фразу, а подстановочные знаки и усечение корня захватывают варианты словоформ.

Посмотрите на пример из рекрутинга: ('Java Developer' OR 'Backend Engineer') AND (Spring OR Spring Boot) NOT 'Junior'. Этот запрос точно определяет, что должно быть найдено, и вы понимаете, почему каждый результат был возвращён.

Сильные стороны булева поиска — контроль и прозрачность. Вы сами определяете область поиска и можете предсказуемо расширять или сужать её.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Естественно-языковой поиск: интерпретация намерения

Естественно-языковой поиск позволяет сформулировать потребность обычными словами — как вопрос коллеге. Современные версии такого поиска являются семантическими: они математически представляют значение текста.

В рекрутинге это выглядит так: «Find a senior Java developer with experience in fintech and AWS, based in London». Без единого оператора система интерпретирует намерение и автоматически строит умные фильтры, навыки и синонимы.

Исследования показывают, что естественно-языковой поиск как минимум не уступает булевому по качеству результатов.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Естественно-языковой поиск: интерпретация намерения

Сильные стороны — охват и лёгкость использования. Можно исследовать незнакомую область, не зная точных терминов. Слабость — контроль: система ранжирует по своему суждению о релевантности, которое обычно хорошо, но не прозрачно. Вы не можете быть полностью уверены, что система не пропустила что-то, что точный запрос поймал бы. Естественно-языковой поиск превосходен для обнаружения и ориентации, но менее надёжен, когда требуется исчерпывающая полнота.

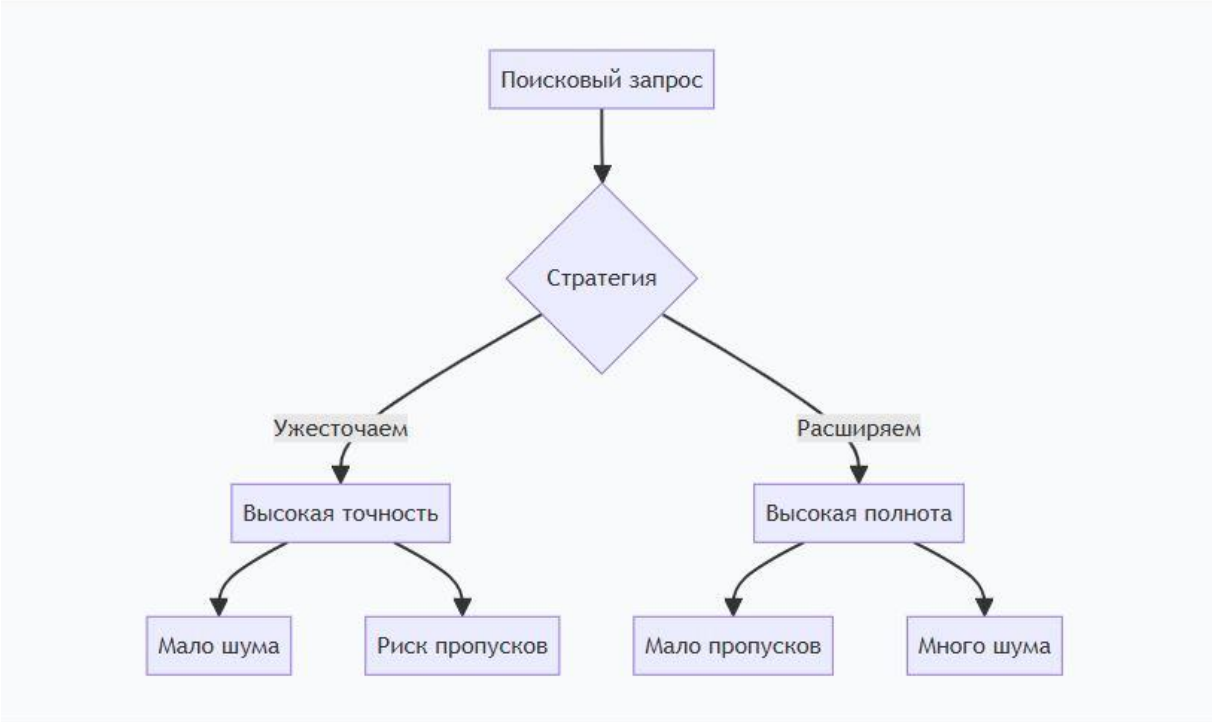
Примерный сценарий: вы ищете информацию о том, как ухаживать за комнатными растениями. Вводите запрос «как поливать фикус», а система находит статью, где написано «увлажнение почвы для фикуса» — потому что понимает смысл, а не только слова.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Естественно-языковой поиск: интерпретация намерения

Естественно-языковой поиск прощает неточные формулировки и находит концептуально связанные материалы даже при использовании другого словаря — это идеальный инструмент для разведки в незнакомой области.

Семантический поиск способен находить документы, в которых нет ни одного слова из вашего запроса, если смысл совпадает.




Точность и полнота тянут в противоположные стороны: ужесточение запроса снижает полноту, расширение — точность.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Комбинированная стратегия: как ищут профессионалы

Опытные исследователи не выбирают между булевым и естественно-языковым поиском — они используют оба в связке. Каждый метод покрывает пробел, который другой оставляет: естественно-языковой поиск находит концептуально связанные материалы с другим словарём, булев — гарантирует, что все документы, соответствующие точному паттерну, будут найдены.

Процедура из шести шагов: 1) начать с широкого вопроса на простом языке; 2) прочитать лучшее решение или мнение; 3) извлечь термины; 4) построить точный булев запрос с этими терминами; 5) запустить булев поиск для исчерпывающего сбора; 6) сравнить результаты обоих проходов.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Комбинированная стратегия: как ищут профессионалы

Такой подход признаёт, что выбор не является либо-либо: сильные стороны одного — слабости другого. Начинайте с естественно-языкового поиска для разведки и обнаружения терминологии, затем переключайтесь на булев для точного и исчерпывающего сбора, когда правильные термины уже известны.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Типы запросов: что стоит за словами

Понимание типов запросов — важный элемент формулирования, потому что каждый тип требует своего подхода. Информационные запросы — самый распространённый тип: пользователь вводит фразу, чтобы что-то узнать, и в выдаче появляются блоги, форумы, видеоуроки. Например, «как выбрать пылесос для квартиры» — человек хочет узнать критерии выбора.

Навигационные запросы возникают, когда человек хочет попасть на конкретный сайт: он уже знает название бренда или компании, но вводит его в поисковую строку, а не в адресную. «allo интернет-магазин» — пользователь хочет попасть на конкретный сайт, а не читать обзоры.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Типы запросов: что стоит за словами

Коммерческие запросы — самые ценные для интернет-магазинов: человек уже хочет купить и вводит фразы с маркерами «купить», «заказать», «доставка», «цена», а также бренды, модели, характеристики, локации.

Существует также коммерчески-исследовательский тип: пользователь ещё не выбрал товар, но активно сравнивает варианты. Высокая детализированность запроса указывает на готовность платить: длинные запросы (long-tail) часто приносят больше всего продаж, хотя и имеют низкую частотность. Каждый тип требует своего контента: информационные — статей, навигационные — оптимизированной главной страницы, коммерческие — карточек товаров, исследовательские — сравнительного контента.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

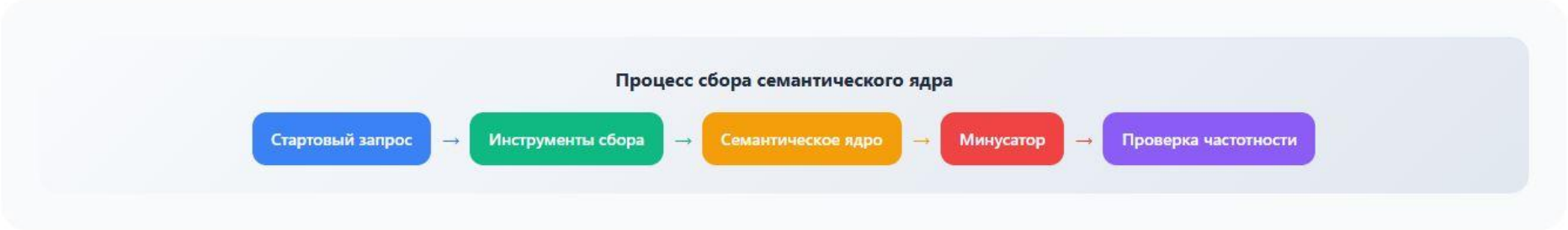
## Типы запросов: что стоит за словами

| ТИП ЗАПРОСА | НАМЕРЕНИЕ ПОЛЬЗОВАТЕЛЯ | ПРИМЕР |
|---|---|---|
| **Информационный** | Что-то узнать | как выбрать пылесос для квартиры |
| **Навигационный** | Попасть на конкретный сайт | allo интернет-магазин |
| **Коммерческий** | Купить товар | купить ноутбук HP |
| **Длинный (long-tail)** | Очень конкретная покупка | приобрести серебристый iPhone 16 256 ГБ |

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Подбор ключевых слов: опора на данные

### Семантическое ядро: как собирают ключевые слова




Подбор ключевых слов — это не произвольный выбор, а процесс, опирающийся на данные о реальных запросах пользователей.

Подбор ключевых слов — это процесс сбора семантического ядра: набора запросов, которые аудитория реально использует в поисковиках. Выбор терминов не должен быть произвольным — он должен опираться на данные о том, как реальные пользователи формулируют свои информационные потребности.

# ФОРМУЛИРОВАНИЕ ПОИСКОВЫХ ЗАПРОСОВ

## Выводы

**Проблема словаря** — главный вызов формулирования: один и тот же концепт может описываться десятками терминов, и поиск по одной формулировке пропускает большинство релевантных документов.

**Булев поиск** даёт контроль и воспроизводимость, но требует заранее предугадать все возможные формулировки.

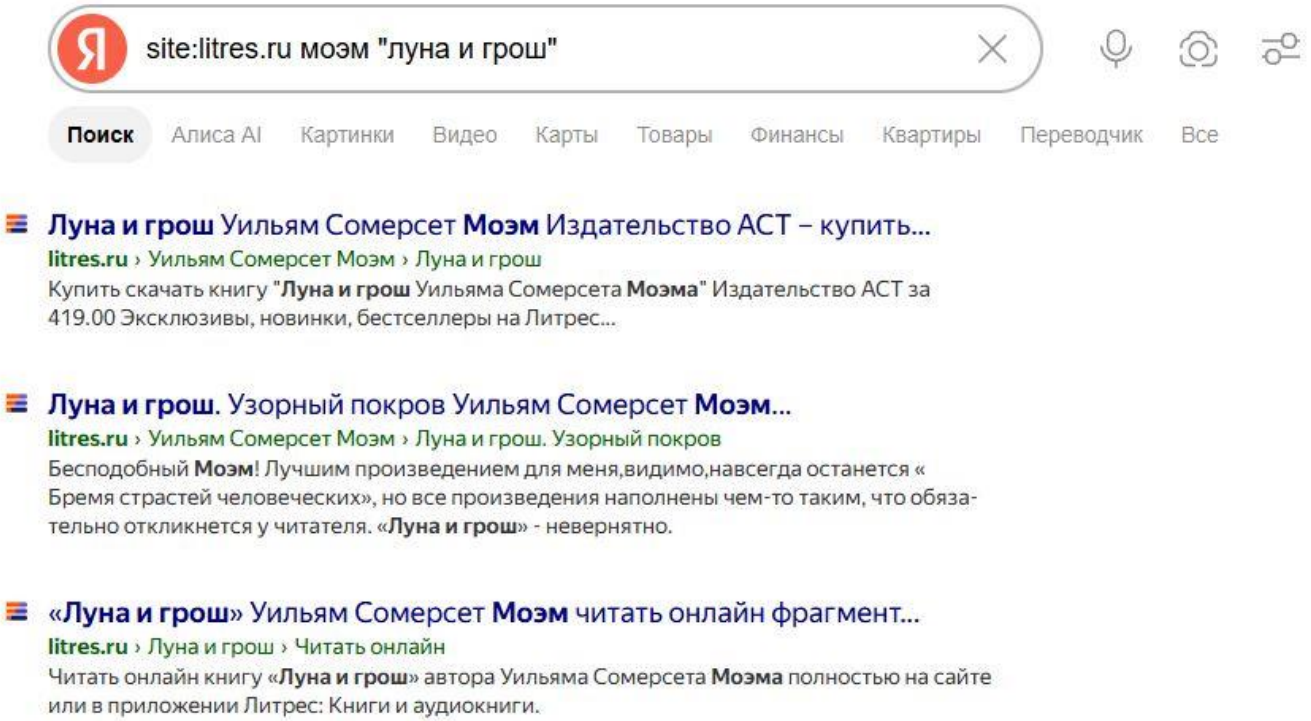
**Естественно-языковой поиск** интерпретирует намерение и находит синонимы, но его ранжирование непрозрачно.

**Точность и полнота** тянут в противоположные стороны: ужесточение запроса снижает полноту, расширение — точность.

**Профессионалы комбинируют оба подхода**: естественно-языковой для разведки, булев для исчерпывающего сбора.

**Типы запросов** (информационные, навигационные, коммерческие) требуют разных стратегий формулирования.

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА



# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Что такое операторы и почему они работают по-разному

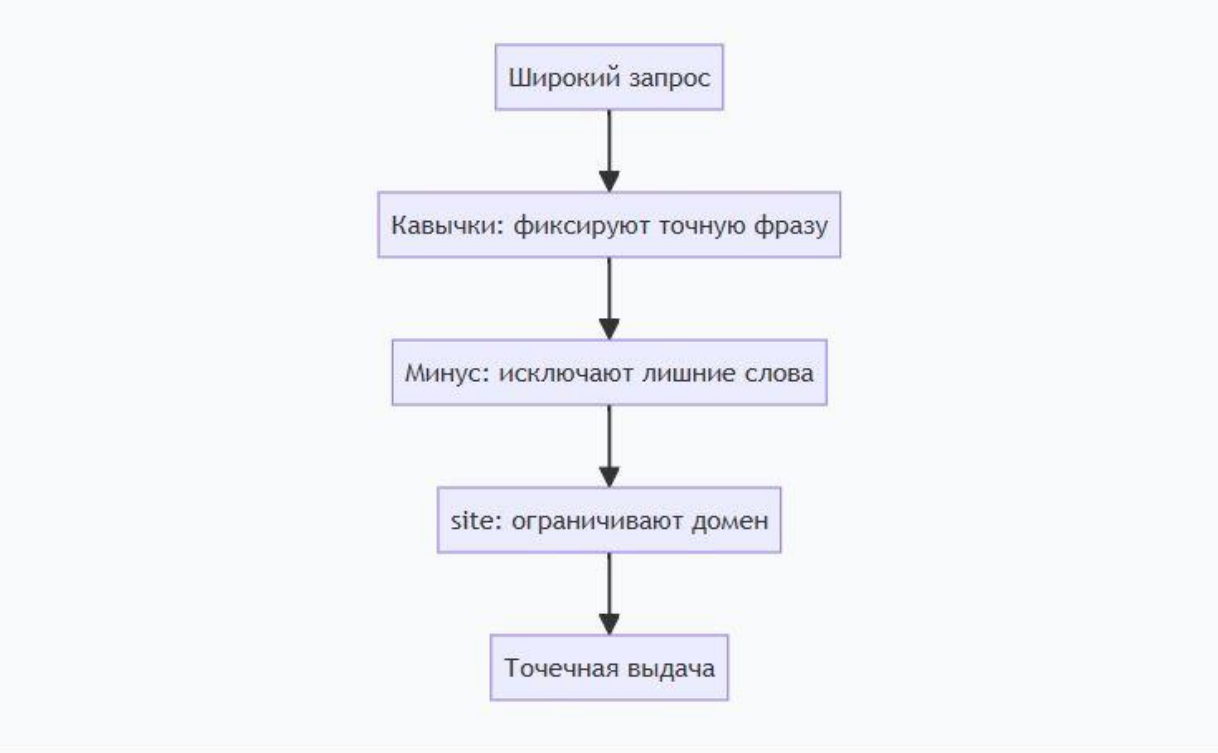
Поисковые операторы — это специальные символы, команды и модификаторы запросов, которые позволяют управлять алгоритмами поисковых систем для получения более точных и релевантных результатов. Они работают как фильтры и инструкции, указывающие поисковой машине, как именно интерпретировать введённый запрос.

Основная задача операторов — дать пользователю больше контроля над процессом поиска и значительно повысить точность получаемых результатов. Современные алгоритмы Google и Яндекса поддерживают десятки различных операторов, каждый из которых решает конкретные задачи: от простого исключения ненужных слов до сложного поиска файлов определённого типа на конкретных доменах.

Но есть важный нюанс: синтаксис у Google и Яндекса различается. Совпадают только кавычки, минус и site — а заголовок, тип файла и логическое «или» пишутся по-разному. Чужой оператор Яндекс молча игнорирует: ошибки не покажет, просто выдаст результаты по остальным словам. Так и рождается ощущение, что «операторы не работают».

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Как операторы сужают выдачу




Каждый оператор — это фильтр, который отсекает часть выдачи. Чем больше условий, тем уже результат.

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Кавычки: «я знаю, что ищу, не помогай»

Кавычки — первый и самый простой оператор. Он заставляет поисковую систему искать точное совпадение слов в указанном порядке. Запрос, заключённый в кавычки, возвращает страницы, на которых этот запрос содержится в таком виде, как он указан в строке поиска — фиксируется количество слов во фразе.

Без кавычек поисковая система может менять порядок слов или вставлять между ними другие слова. Кавычки фиксируют и порядок, и состав фразы — это критически важно для профессионального поиска информации.

Этот оператор особенно полезен для поиска цитат, названий или специфических терминов. В контексте продвинутого поиска кавычки — единственный надёжный способ сказать алгоритму: «я знаю, что ищу, не помогай».

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Минус и плюс: убрать лишнее, закрепить нужное

Оператор минус исключает из результатов страницы, содержащие указанное слово. Если перед словом поставить знак минуса, из выдачи будут исключены все страницы, на которых это слово встречается. Важно ставить минус без пробела перед исключаемым словом — иначе поисковик воспримет его как обычный символ. Например, запрос Apple -iPhone покажет информацию о компании Apple, исключив упоминания iPhone.

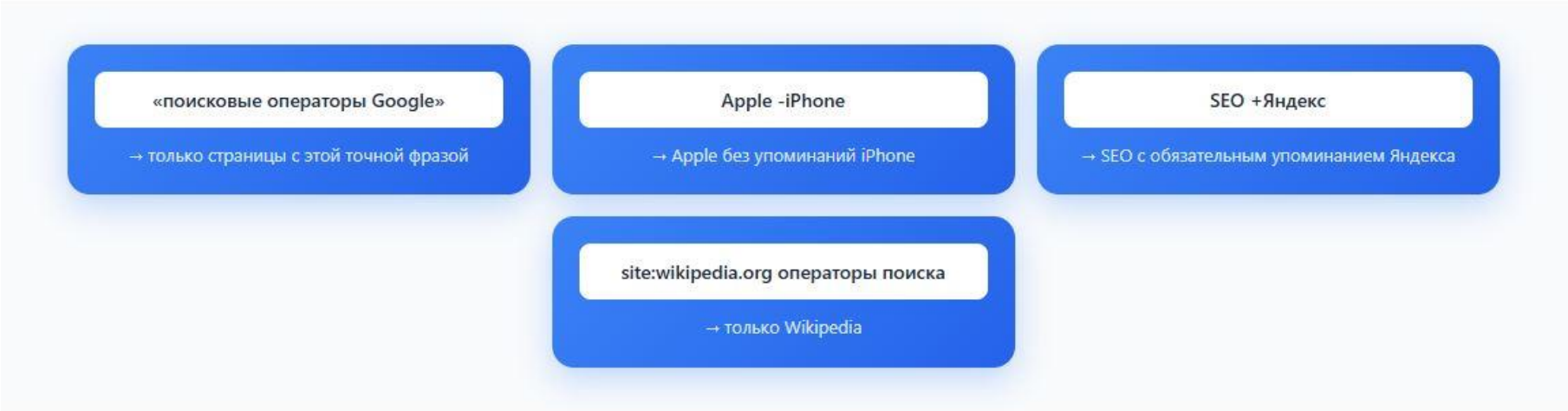
Минус особенно полезен при поиске многозначных терминов, когда нужно исключить один из смыслов. Он работает и с операторами: -site:pinterest.com исключит целый домен из результатов. Для исключения нескольких терминов можно использовать несколько минусов.

Знак плюса — ещё один уточняющий оператор, который размещается перед словом. Плюс в основном используется для фиксации стоп-слов — служебных частей речи, местоимений и так далее. По умолчанию они могут игнорироваться поисковой системой, но если их наличие влияет на смысл фразы, используйте оператор, чтобы получать релевантную выдачу. Пример: SEO +Яндекс найдёт результаты по SEO, где обязательно упоминается Яндекс.

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Минус и плюс: убрать лишнее, закрепить нужное

### Операторы в действии: примеры запросов




Один и тот же запрос с разными операторами даёт принципиально разную выдачу.

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Звёздочка и OR: подставить неизвестное, расширить поиск

Оператор звёздочка позволяет подставить во фразу любые пропущенные слова. Звёздочка может заменять любое количество слов, которые находятся между введёнными фразами или словами. Например, запрос маркетинг в * сетях может найти «маркетинг в социальных сетях», «маркетинг в рекламных сетях» и так далее. Это особенно удобно, когда вы не знаете точную формулировку: звёздочка работает как заполнитель для любого слова при поиске вариаций фраз.

Оператор вертикальная черта используется для группировки слов при сложных запросах и обозначает логическое «ИЛИ». Он позволяет получить выдачу страниц, которые содержат любое слово из запроса — или одно, или другое. Аналогом оператора OR является команда | — они работают одинаково, поэтому можно использовать любой из них.

OR следует записывать в верхнем регистре, чтобы поисковая система не принимала его за обычное ключевое слово. Пример: обувь redwood | baldinini | fabi найдёт страницы, где упоминается обувь и хотя бы один из указанных брендов.

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Как работают операторы домена, заголовка и URL

| ОПЕРАТОР | ЧТО ДЕЛАЕТ | ПРИМЕР |
|---|---|---|
| **site:** | Ограничивает поиск конкретным доменом или поддоменом | site:wikipedia.org операторы поиска |
| **intitle:** | Ищет слово в заголовках страниц (тег title) | intitle:SEO |
| **allintitle:** | Требует все слова фразы в заголовке | allintitle:marketing strategy guide |
| **inurl:** | Ищет страницы, в URL которых есть запрос | inurl:blog |
| **related:** | Находит сайты, связанные с вашим | related:raiffeisen.ru |

Разница между intitle: и allintitle: — в количестве слов. intitle: требует только одно слово в заголовке, а allintitle: — все слова. Например, по запросу intitle:marketing strategy возвращается более 27 200 000 результатов, а allintitle:marketing strategy guide — только 89 000. Эта разница наглядно показывает, как оператор сужает выдачу.

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Правила синтаксиса: как не сломать запрос

| № | Требование |
|---|---|
| 1 | Не ставьте пробел после двоеточия |
| 2 | Минус ставьте без пробела |
| 3 | OR пишите в верхнем регистре |
| 4 | Звёздочку используйте только внутри кавычек |
| 5 | Отминусовывайте цифры через кавычки |
| 6 | Помните про сужение выборки |

# БАЗОВЫЕ ОПЕРАТОРЫ ПОИСКА

## Базовый поиск

| № | Действие |
|---|---|
| 1 | Сформулируйте запрос без операторов и посмотрите на выдачу |
| 2 | Добавьте кавычки для точной фразы, если нужно |
| 3 | Исключите лишние слова через минус без пробела перед словом |
| 4 | Ограничьте домен через site: например site:example.com |
| 5 | Проверьте, что между операторами и словами нет лишних пробелов |

# ПРОДВИНУТЫЕ ОПЕРАТОРЫ И ФИЛЬТРЫ

**Кавычки («» или «») Восклицательный знак (!) Плюс (+) Минус (-) Звёздочка (*) Квадратные скобки ([]) Круглые скобки (()) вертикальная черта (|)**

**site: url: host: domain: title: mime: lang: date:**

# ПРОДВИНУТЫЕ ОПЕРАТОРЫ И ФИЛЬТРЫ

## filetype: — поиск по формату документа

Оператор filetype: — это команда, которая говорит поисковику: «покажи только документы определённого формата». Он работает в обеих крупных системах — Google и Bing — и понимает такие форматы, как PDF, DOC, XLS, PPT и TXT. Яндекс также понимает этот оператор.

Как это выглядит на практике? Запрос photoelektrischer effekt artikel filetype:pdf вернёт только PDF-документы по теме фотоэлектрического эффекта — никаких обычных страниц, только файлы. Это незаменимо, когда вы ищете научные статьи, техническую документацию или исследовательские отчёты, которые чаще всего публикуются именно в PDF.

Вы готовите реферат и ищете свежие исследования. Вместо того чтобы листать блоги, вы вводите “квантовая запутанность filetype:pdf” и сразу получаете PDF-файлы с научными статьями, отсекая всё лишнее.

# ПРОДВИНУТЫЕ ОПЕРАТОРЫ И ФИЛЬТРЫ

## before: и after: — фильтрация по дате обновления

Операторы before: и after: ограничивают результаты документами, которые были обновлены до или после указанной даты.

Самое интересное начинается, когда вы комбинируете оба оператора. Запрос avengers endgame after:2019-03-01 before:2019-03-05 найдёт документы, обновлённые в период с 1 по 5 марта 2019 года — то есть в узком временном окне. Это мощный инструмент, когда нужно найти свежие новости или материалы, появившиеся в конкретный промежуток времени.




after: задаёт нижнюю границу, before: — верхнюю. Вместе они образуют временное окно поиска.

# ПРОДВИНУТЫЕ ОПЕРАТОРЫ И ФИЛЬТРЫ

## define: и cache: — определения и кэш

Оператор define: предназначен для поиска определения слова или термина. Запрос define:epistemology в качестве первого результата покажет определение термина «эпистемология». В отличие от filetype: или site:, define: не требует дополнительного ключевого слова — достаточно ввести сам оператор и интересующий термин. Это быстрый способ получить авторитетное определение, не переходя на сторонние словарные сайты.

Оператор cache: показывает сохранённую поисковой системой версию веб-страницы. Для SEO-специалиста это ценный диагностический инструмент: он позволяет увидеть, как поисковик «видит» страницу и какой контент считает релевантным для извлечения. Например, запрос cache:https://www.bruceclay.com/seo/ покажет кэшированную версию указанной страницы.

Но здесь есть важный нюанс: cache: — оператор ненадёжный. В актуальных руководствах его относят к категории «недостоверных» или даже «отключённых». Он не был официально отменён Google, но результаты его работы непостоянны. Поэтому при использовании cache: будьте готовы к тому, что он может не сработать, и держите в запасе альтернативные методы проверки кэша.

# ПРОДВИНУТЫЕ ОПЕРАТОРЫ И ФИЛЬТРЫ

## Графические фильтры: когда не нужно помнить синтаксис

Не все инструменты требуют ручного ввода команд. Поисковые системы предоставляют графические фильтры, которые делают то же самое, но через интерфейс. В Google после выполнения поиска под строкой запроса появляются фильтры, позволяющие ограничить результаты по типу контента. Важная особенность: точный набор фильтров и их порядок динамичны — они зависят от того, что системы Google считают наиболее полезным для конкретного запроса. Если нужный фильтр не отображается, воспользуйтесь опцией «Все», чтобы просмотреть доступные.

В Bing есть страница расширенного поиска, доступная по ссылке «Advanced search» внизу страницы результатов. Там можно задать язык и регион, свежесть результатов, настройки безопасного поиска, ограничения по домену, фильтры по типу файлов и поиск по конкретному сайту. Преимущество графического интерфейса в том, что Bing сам строит правильные операторы за кулисами — это удобно, если вы забыли синтаксис. А заодно можно подсмотреть, какие операторы были созданы, и научиться использовать их вручную в следующий раз.

# ПРОДВИНУТЫЕ ОПЕРАТОРЫ И ФИЛЬТРЫ

## Комбинирование операторов: от простого к сложному

| | |
|---|---|
| 1 | Сформулируйте главную тему запроса одним словом или короткой фразой |
| 2 | Добавьте операторы для сужения: intitle:, inurl:, site: |
| 3 | Исключите ненужные сайты с помощью минус-оператора |
| 4 | Проверьте, что после двоеточий нет пробелов |
| 5 | Если результатов слишком много, добавьте еще операторы; если мало – уберите лишние |

# ПРОДВИНУТЫЕ ОПЕРАТОРЫ И ФИЛЬТРЫ

## Надёжность операторов: работающие, ненадёжные, отключённые

Продвинутые операторы — это не вечный стандарт. В современных руководствах их делят на три категории: работающие, ненадёжные и отключённые. К ненадёжным относятся операторы, которые Google официально не отменял, но которые дают непостоянные результаты — например, cache: и операторы для поиска по датам, требующие юлианских дат.

Отключённые операторы — это те, которые Google официально прекратил поддерживать. Например, оператор ~ (тильда) для включения синонимов был отключён в 2013 году.

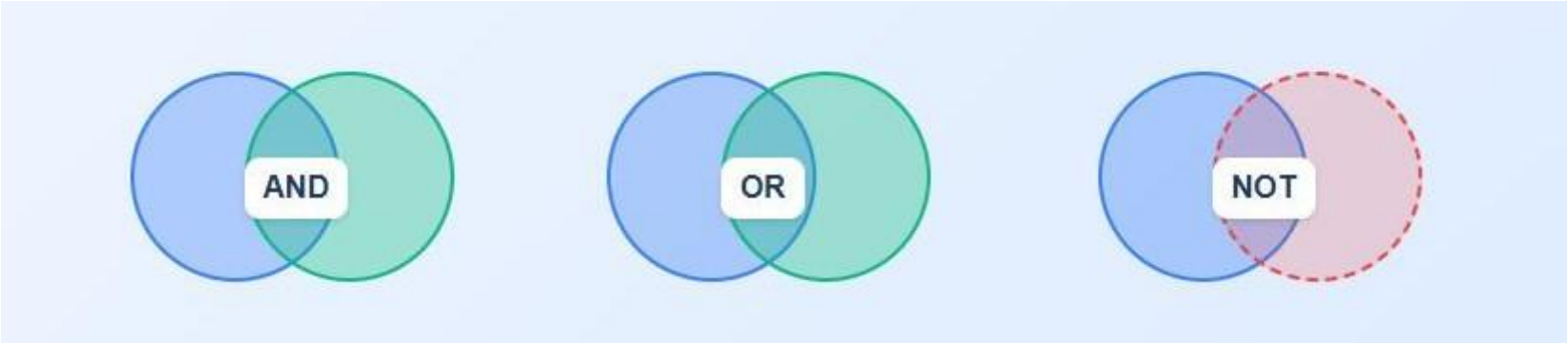
Почему это происходит? Алгоритм поисковика становится умнее, и потребность в ручных операторах снижается. То, что раньше требовало явной команды, теперь поисковик понимает сам. Поэтому при работе с продвинутыми операторами важно проверять актуальность: то, что работало вчера, может перестать работать завтра.

# ПРОДВИНУТЫЕ ОПЕРАТОРЫ И ФИЛЬТРЫ

## Ключевые выводы

| |
|---|
| filetype: ограничивает поиск документами определённого формата (PDF, DOC, XLS, PPT, TXT) и работает в Google и Bing |
| before: и after: фильтруют результаты по дате обновления; их комбинация создаёт временное окно поиска |
| define: мгновенно выдаёт определение термина, а cache: показывает кэшированную версию страницы, но относится к ненадёжным операторам |
| Графические фильтры Google и Bing, включая расширенный поиск, заменяют ручной ввод операторов |
| Операторы делятся на работающие, ненадёжные и отключённые — надёжность нужно проверять |

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ



# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## Поисковая строка — это уравнение, а не список слов

Вы когда-нибудь получали в выдаче тысячи страниц, среди которых не было ни одной нужной? Скорее всего, вы забыли сказать поисковику, что именно вам нужно. Логические операторы AND, OR, NOT — это способ точно объяснить системе, какие слова обязательны, какие желательны, а какие вообще не нужны.

Булева логика, на которой основаны эти операторы, существовала задолго до появления интернета. В русскоязычной традиции операторы называются «И» (AND), «ИЛИ» (OR), «НЕ» (NOT). Освоив их, вы перестанете гадать, что найдёт поисковик, и начнёте управлять выдачей.

Практический пример из Scopus: Запрос unicorn AND maiden сокращает результаты с огромного множества до 1 386 записей, содержащих оба термина.

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## AND: чем больше условий, тем точнее результат

Оператор AND (логическое «И») требует, чтобы в результатах присутствовали **все** указанные термины. Каждый добавленный AND отсекает документы, где отсутствует хотя бы одно слово, поэтому поиск сужается, а количество результатов уменьшается.

Например, запрос college students test anxiety в базе данных будет интерпретирован как college AND students AND test AND anxiety — система найдёт документы, где встречаются все четыре слова. Но обратите внимание: слова могут быть разбросаны по разным частям документа и не связаны друг с другом. Если нужна точная фраза, AND стоит комбинировать с кавычками: "college students" AND "test anxiety".

Во многих системах AND подразумевается по умолчанию. В Google пробел между словами автоматически интерпретируется как AND, поэтому явно писать его не нужно.

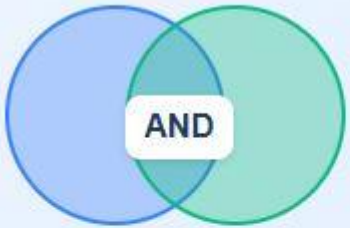
# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## Сравнение

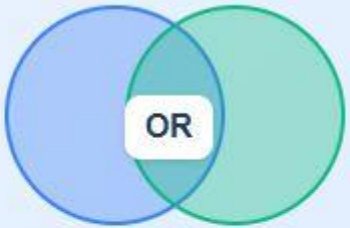
| Поиск с кавычками | | Поиск без кавычек |
|---|---|---|
| Ищет точную фразу | **Точность совпадения** | Ищет отдельные слова |
| Низкая — только заданный порядок слов | **Гибкость** | Высокая — слова могут быть в любом месте |
| Меньше — только точное совпадение | **Риск ложных результатов** | Больше — могут попасться документы с разными сочетаниями |

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

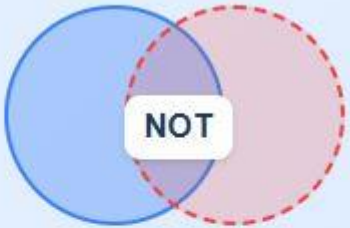
## Логика на кругах: AND, OR, NOT




Пересечение: оба термина обязательны




Объединение: любой из терминов




Исключение: термин не должен встречаться

Три базовых логических оператора: AND — пересечение, OR — объединение, NOT — исключение.

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## OR: когда синонимы — ваши союзники

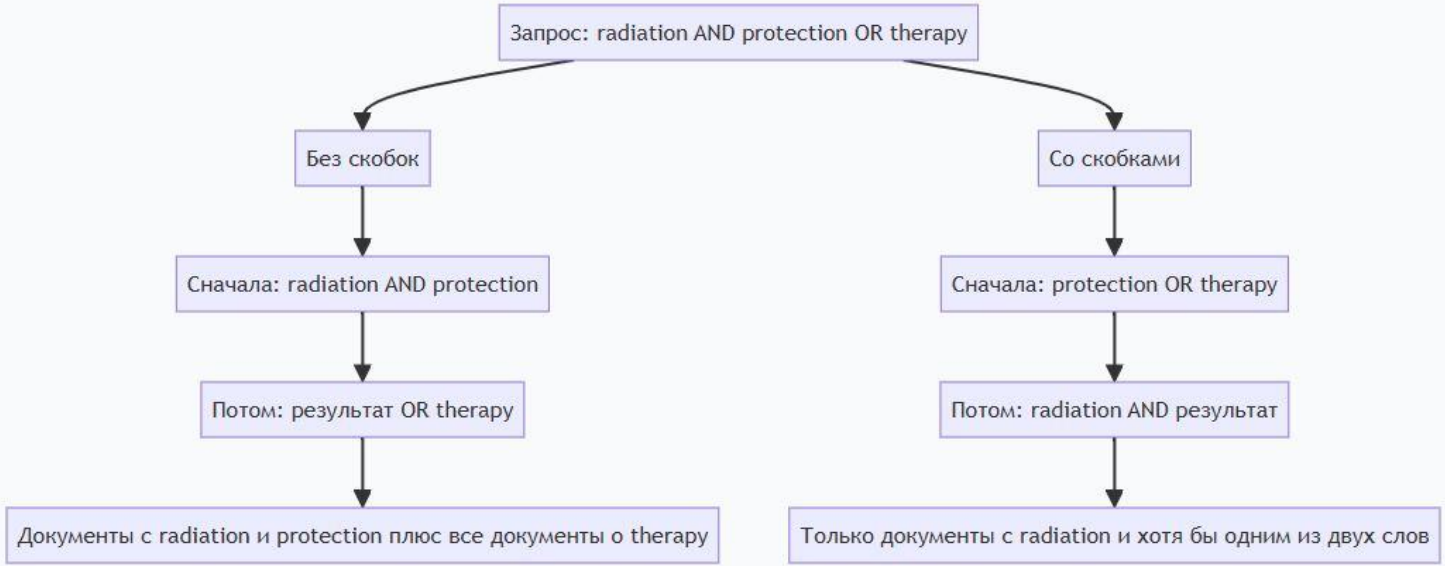
Оператор OR (логическое «ИЛИ») работает противоположно AND: он расширяет поиск, сообщая базе данных, что **любой** из терминов может присутствовать в результатах. Вопреки распространённому заблуждению, OR — это не строгое «или/или». Технически он интерпретируется как «как минимум один термин обязателен, но могут быть возвращены и несколько, и все». Результаты, содержащие несколько терминов, обычно ранжируются выше как более релевантные.

OR особенно полезен при работе с синонимами и альтернативными написаниями. Например, запрос preschool OR "nursery school" позволит найти материалы, где авторы используют разные термины для одного понятия. В Google оператор OR пишется заглавными буквами, также поддерживается символ |.

В Academic Search Ultimate запрос art OR music даёт 1 785 300 результатов, тогда как art AND music — только 76 192. Разница наглядно показывает, как операторы управляют объёмом выдачи.

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## Приоритет операторов: кто выполняется первым?




Большинство баз данных обрабатывают AND раньше OR. Без скобок запрос radiation AND protection OR therapy выполняется как (radiation AND protection) OR therapy, что даёт много лишних документов о therapy.

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## NOT: отсекаем лишнее

Оператор NOT (логическое «НЕ») сужает поиск, исключая документы, содержащие указанный термин. Это эффективный инструмент для удаления ложных результатов — когда запрос возвращает материалы, формально содержащие искомые слова, но не соответствующие реальной цели поиска.

Например, при поиске IT-проектного менеджера запрос "project manager" NOT construction исключит все резюме, где упоминается строительство. NOT особенно полезен, когда ключевое слово имеет несколько значений: например, «Тесла» — это и фамилия учёного, и марка автомобиля.

В Google и многих других системах аналогом NOT служит знак минуса -, который ставится непосредственно перед исключаемым словом без пробела.

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## Собираем сложный запрос по частям

```
(unicorn OR Pegasus) AND maiden NOT tapestry
```

Порядок выполнения:

1. `(unicorn OR Pegasus)` — любое из двух слов
2. `AND maiden` — maiden обязателен
3. `NOT tapestry` — исключить tapestry

Комбинирование всех трёх операторов в одном запросе: скобки группируют OR-выражение, AND связывает обязательный термин, NOT исключает нежелательный.

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## Сценарий

**Поиск рецепта без орехов**

Вы ищете рецепт торта, который может быть с шоколадом или карамелью, но без орехов. Вы вводите: (шоколад OR карамель) AND торт NOT орехи. Система возвращает только те рецепты, где есть шоколад или карамель, обязательно торт, и ни одного упоминания орехов. Так вы быстро находите подходящий вариант.

Перед сложным поиском в незнакомой базе данных всегда загляните в раздел «Справка» или «Помощь» — там указано, как система обрабатывает пробелы и операторы по умолчанию. Это сэкономит вам время и избавит от неожиданных результатов.

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## Пять шагов к сложному запросу

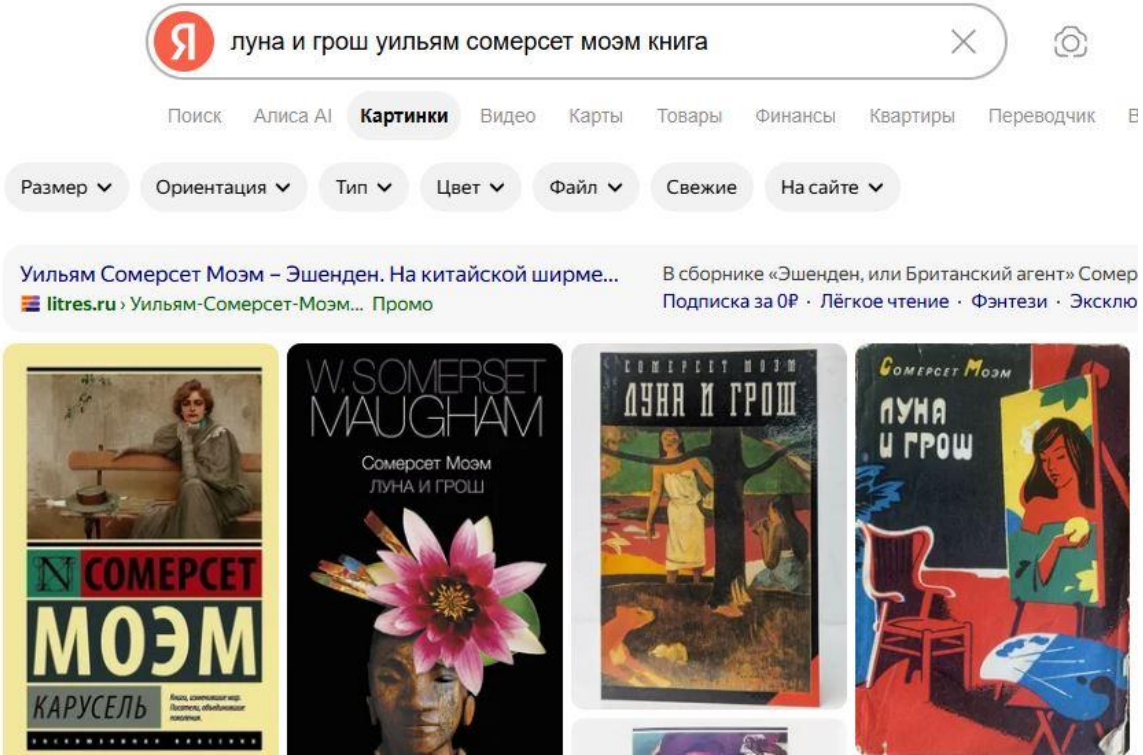
| | | |
|---|---|---|
| 1 | **Определите ключевые слова** | Например, при поиске информации о клонировании: cloning, sheep, human. Запишите все слова, которые могут встретиться в нужных документах. |
| 2 | **Решите, какие термины обязательны** | AND для обязательных, OR для альтернативных, NOT для исключаемых. Спросите себя: без какого слова документ точно не подойдёт? |
| 3 | **Сгруппируйте OR-выражения в скобки** | cloning AND (sheep OR human) — этот запрос найдёт информацию о клонировании овец и о клонировании человека. Скобки гарантируют, что OR выполнится первым. |
| 4 | **Проверьте, не нужно ли исключить лишнее** | Добавьте NOT для удаления нежелательных результатов. Например, если вы ищете клонирование, но не генную инженерию растений, добавьте NOT plants. |
| 5 | **Протестируйте запрос и скорректируйте** | Метод проб и ошибок — лучший способ получить нужные результаты. Посмотрите на первые 10 результатов: если есть лишнее, уточните запрос. |

# ЛОГИЧЕСКИЕ ОПЕРАТОРЫ И КОМБИНИРОВАНИЕ

## Ключевые выводы

| |
|---|
| AND сужает поиск, требуя присутствия всех терминов; в Google он подразумевается по умолчанию между словами. |
| OR расширяет поиск, допуская любой из терминов; особенно полезен для синонимов и альтернативных написаний. |
| NOT исключает нежелательные термины и помогает убирать ложные результаты; в разных системах обозначается по-разному. |
| Скобки группируют условия и задают порядок выполнения; большинство баз данных обрабатывают AND раньше OR. |
| Разные системы по-разному реализуют логические операторы: от автоматического AND в Google до явных операторов в научных базах и русских «или»/«и» на hh.ru. |

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ



# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

## Вертикальный поиск: когда общий поиск бессилен

Представьте: вы ищете вакансию Java-разработчика. Вводите запрос в Google и получаете смесь из новостей, статей, видео и рекламы. Где-то среди этого хаоса затерялись реальные вакансии — но чтобы их найти, придётся перелопатить десятки страниц.

А теперь представьте, что вы сразу попадаете на сайт, где собраны только вакансии. Никакого шума. Именно так работает вертикальный поиск.

Ключевое архитектурное отличие — в способе индексации. Общие поисковики используют веб-краулер, который обходит весь интернет. Вертикальные системы обычно применяют фокусированный краулер: он индексирует только релевантные страницы по заранее определённой теме. Это фундаментально меняет и скорость, и качество результатов.

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

## Что такое вертикальный поиск

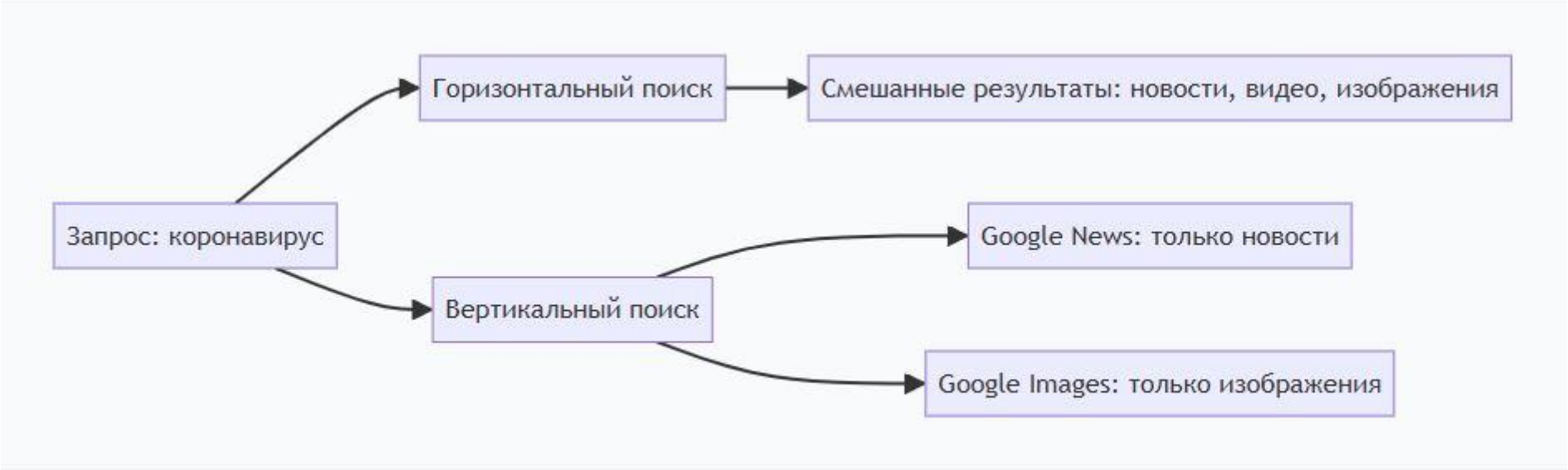
Вертикальный поиск (vertical search) — это тип поисковой системы, которая фокусируется на конкретном сегменте онлайн-контента. В отличие от универсальных поисковиков, которые пытаются индексировать значительные части Всемирной паутины, вертикальные системы работают с узкой областью. Их также называют специализированными (specialty) или тематическими (topical) поисковыми системами.

Что может быть такой областью? Практически что угодно: тематика (медицина, право, недвижимость), тип медиа (изображения, видео, аудио) или жанр контента (новости, вакансии, научные статьи). Например, Zillow — вертикальная поисковая система для недвижимости, Kayak — для путешествий, LinkedIn — для профессиональных связей, а Google Maps — для карт и навигации.

Вертикальный поиск — это не отдельная технология, а подход. Некоторые вертикальные сайты фокусируются на одной вертикали, другие включают несколько вертикальных поисков в одну систему.

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

## Горизонтальный vs вертикальный поиск




Один и тот же запрос даёт принципиально разные результаты в горизонтальном и вертикальном поиске.

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

## Почему вертикальный поиск выигрывает

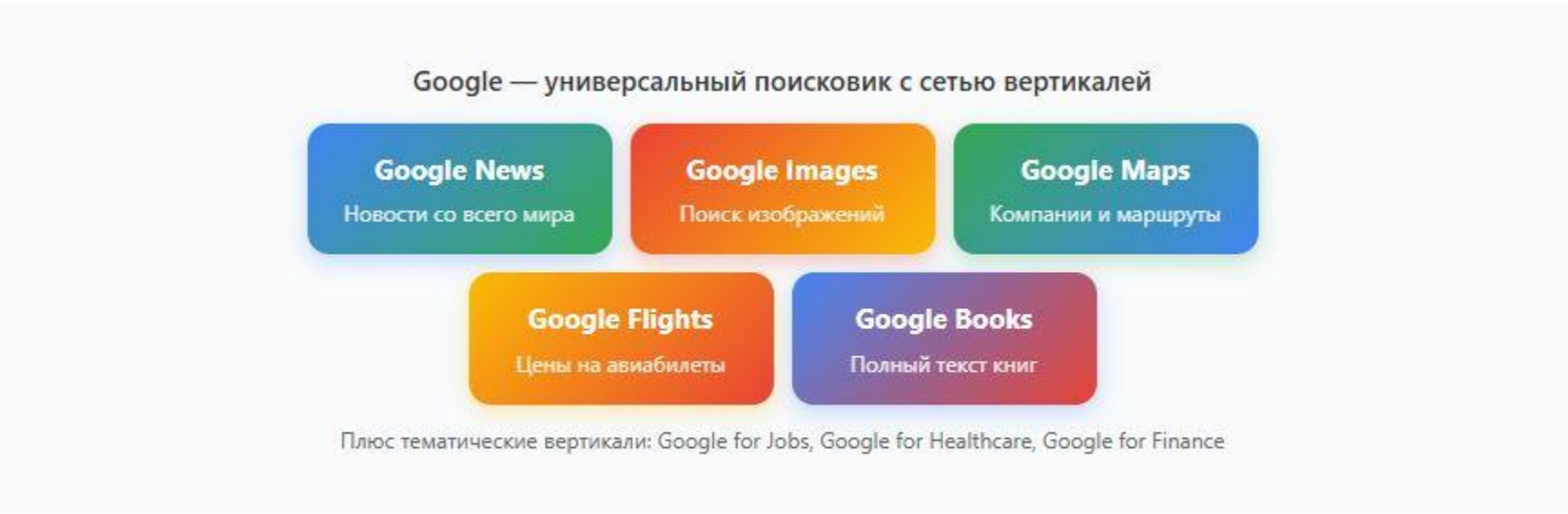
Зачем вообще использовать вертикальный поиск, если Google и так умеет находить всё? Ответ прост: для нишевых задач он работает заметно лучше.

Три ключевых преимущества вертикального поиска. Первое — целевые результаты: фокусируясь на конкретной отрасли или типе контента, вертикальный поиск отсекает всё лишнее. Второе — скорость и эффективность: вы не тратите время на просмотр нерелевантных типов контента. Третье — качество информации: благодаря фокусировке на конкретной области, данные в вертикальном поиске, как правило, превосходят то, что выдают универсальные системы.

Пользователи, ищущие нишевую тему, получают явную выгоду: стандартный поисковик возвращал бы слишком широкие результаты, и большая часть информации не была бы релевантна конкретному запросу.

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

## Вертикальные поисковые системы Google




Эти инструменты работают на базовой поисковой технологии Google, но настроены под конкретные типы контента.

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

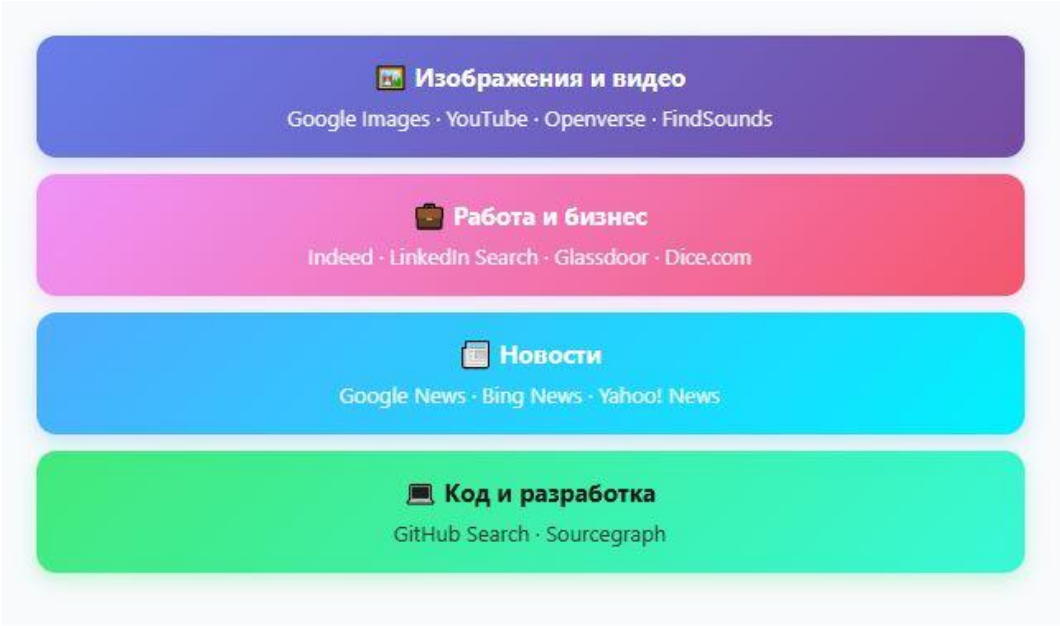
## Категории вертикальных поисковых систем

Вертикальные поисковые системы покрывают самые разные категории контента. В области изображений, видео и медиа лидирует Google Images — ведущая система поиска изображений с интегрированным обратным поиском. YouTube — крупнейшая видеопоисковая система, которую обычно считают второй по использованию поисковой системой в мире. Для открытого контента есть Openverse: он ищет изображения и аудио с лицензиями Creative Commons и поддерживается проектом WordPress. A FindSounds специализируется на звуковых эффектах и аудиоклипах.

Перед поиском определите тип нужного контента и сразу открывайте соответствующую вертикаль — так вы сэкономите время и получите более релевантные результаты.

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

## Категории вертикальных поисковых систем




Четыре основные категории вертикального поиска и их ключевые представители.

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

## Шаги использования вертикальных поисковых систем

| |
|---|
| Сформулируйте, какую именно информацию вы ищете(вакансию, новость, код, изображение) |
| Составьте список вертикальных систем, которые специализируются на этой области |
| Сравните их по качеству результатов и удобству интерфейса |
| Протестируйте каждую на одном и том же запросе |
| Выберите ту, что дает наиболее релевантные результаты, и сохраните ее в закладки |

# ПОИСК В СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЯХ

## Ключевые выводы

| |
|---|
| Вертикальный поиск фокусируется на конкретном сегменте контента — тематике, типе медиа или жанре — в отличие от горизонтального, охватывающего весь интернет. |
| Главные преимущества вертикального поиска — целевые результаты, скорость и более высокое качество информации в рамках специализированной области. |
| Google управляет сетью вертикальных инструментов: Google News, Images, Maps, Flights, Books, а также тематическими вертикалями для здоровья, финансов и работы. |
| Ключевые вызовы вертикального поиска — курирование данных и принятие пользователями; будущее — за интеграцией ИИ и персонализацией результатов. |

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ




Scopus

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Google не знает науки: почему учёные ищут в другом месте

Представьте: вы ищете свежее исследование о лечении диабета. Обычный поисковик выдаст тысячи страниц — от блогов до рекламы БАДов. Но настоящие научные статьи, прошедшие рецензирование, спрятаны в специализированных базах данных, куда Google почти не заглядывает.

Научные базы данных — это не просто «каталоги статей». Это структурированные коллекции проверенной информации: рецензируемые журнальные статьи, диссертации, материалы конференций. Здесь действуют свои правила поиска, свои фильтры и свои показатели значимости.

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Типология научных баз данных

Научные базы данных неоднородны. Чтобы не утонуть в их многообразии, полезно разделить их на несколько категорий по содержанию и функциям.

**Библиографические базы данных** (их также называют тематическими каталогами) содержат собрания публикаций по определённой области знаний. Они хранят библиографические данные — авторов, названия, журналы, годы — и при необходимости информацию о специализированных текстах. **Библиотечные каталоги** (каталоги фондов) — это перечни фондов конкретной библиотеки или библиотечного объединения.

**Полнотекстовые коллекции**, такие как ScienceDirect и ACM, дают доступ к полным текстам публикаций. А **наукометрические базы данных** — это библиографические и реферативные базы, которые служат инструментом для отслеживания цитируемости научных публикаций и формируют статистику, характеризующую динамику показателей востребованности и индексов влияния деятельности учёных и организаций.

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Типы научных баз данных




Классификация научных баз данных по содержанию и функциям.

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Scopus и Web of Science: два гиганта

Это две крупнейшие коммерческие библиографические базы данных, охватывающие научную литературу практически из всех дисциплин.

**Scopus** разработан компанией Elsevier и выпущен в 2004 году. Это крупнейшая в мире коллекция абстрактов и цитирования рецензируемой исследовательской литературы, содержащая более 76 миллионов записей. База включает более 20 500 наименований от более чем 5000 международных издателей и охватывает естественные, технические, медицинские, социальные и гуманитарные науки. Scopus предлагает расширенные функции поиска с фильтрацией по автору, названию, теме, дате публикации и другим критериям.

**Web of Science** разработан Институтом научной информации (ISI) и поддерживается Clarivate Analytics. База включает более 21 000 научных журналов, а также конференции и книги. WoS известен высоким стандартом качества данных и строгими критериями включения публикаций. Он делает акцент на анализе цитирования и библиометрических показателях.

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Как выбрать базу данных

**Рецензируемые статьи**
Scopus или Web of Science

**Широкий охват**
Google Scholar — включая нерецензируемые

**Биомедицина**
PubMed / MEDLINE

**Образование**
ERIC

**Инженерия и IT**
IEEE Xplore

**Гуманитарные науки**
JSTOR

Шпаргалка: какую базу выбрать под конкретную задачу.

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Специализированные базы данных по дисциплинам

**ERIC** (Education Resources Information Center) специализируется на литературе по образовательным наукам и содержит более 1,2 миллиона библиографических записей по образовательным темам. **IEEE Xplore** — ведущая база данных в области инженерии и компьютерных наук, включающая не только журнальные статьи, но и материалы конференций, стандарты и книги.

**JSTOR** — одна из крупнейших многодисциплинарных баз данных, предоставляющая доступ к тысячам академических журналов, книгам и первоисточникам; она особенно полезна для гуманитарных и социальных наук. **ScienceDirect** — платформа Elsevier для научных, технических и медицинских исследований, содержащая более 2500 рецензируемых журналов и более 40 000 книг. **SpringerLink** — платформа Springer Nature, включающая миллионы рецензируемых документов из более чем 4000 журналов и 300 000+ книг.

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Наукометрические показатели

Научные базы данных — это не только поиск статей, но и инструменты оценки их значимости. Наукометрические базы данных предоставляют статистику, характеризующую динамику показателей востребованности и индексов влияния деятельности учёных и организаций.

**Индекс цитирования** — это принятая в научном мире мера «значимости» научной работы учёного или коллектива, определяемая количеством ссылок на публикацию или фамилию автора в других источниках. **Индекс Хирша** (h-index) — количественная характеристика учёного, основанная на количестве его публикаций и количестве цитирований этих публикаций. **Импакт-фактор** (IF) — показатель влияния научного издания.

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Как работает поиск по цитированию




Цитирование как инструмент расширения поиска: от одной статьи к целому пласту литературы.

# ПОИСК В НАУЧНЫХ И АКАДЕМИЧЕСКИХ БАЗАХ ДАННЫХ

## Ключевые выводы

| |
|---|
| Научные базы данных делятся на библиографические, библиотечные каталоги, полнотекстовые коллекции и наукометрические. |
| Scopus и Web of Science — крупнейшие коммерческие базы для рецензируемой литературы с инструментами цитирования. |
| PubMed и MEDLINE — стандарт для биомедицинских наук с контролируемым словарём MeSH. |
| Google Scholar бесплатно индексирует широкий спектр источников, включая нерецензируемые. |
| Доступ к коммерческим базам — через институциональные подписки, VPN и DBIS. |
| Наукометрические показатели (индекс Хирша, импакт-фактор, SJR) помогают оценить влияние публикаций. |

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ



# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Красивый сайт, умные слова — а информация ложная

Представьте: вы открываете сайт с доменом .org, профессиональным дизайном и уверенными формулировками. Автор рассуждает о сложной теме, приводит цифры, делает выводы. Всё выглядит безупречно. Но при проверке оказывается, что данные устарели на десять лет, автор — вымышленное лицо, а цель материала — продать вам сомнительный курс.

Наш мозг эволюционно не приспособлен к оценке цифровых источников. Мы доверяем тому, что выглядит солидно, и легко попадаемся на удочку дезинформации. Хорошая новость: существуют системные методы проверки, которые превращают оценку из интуитивного «вроде надёжно» в осознанный процесс.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## CRAAP-тест: пять вопросов, которые спасают от дезинформации

CRAAP-тест — это мнемонический чек-лист из пяти критериев: **Currency** (актуальность), **Relevance** (релевантность), **Authority** (авторитетность), **Accuracy** (точность) и **Purpose** (цель). Он помогает оценить объективную надёжность источника в академических дисциплинах.

Тест разработала библиотекарь Сара Блейксли весной 2004 года, когда готовила воркшоп для преподавателей курсов. Она не могла запомнить критерии оценки источников и придумала акроним, чтобы дать студентам простой способ определения достоверности.

Важно понимать: CRAAP — это не пропускной барьер «сдал/не сдал». Источник может получить низкую оценку по одному критерию, но остаться ценным. Например, правительственный отчёт 1995 года не отвечает критерию актуальности, но является авторитетным первоисточником по политике той эпохи.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Студент и красивый блог

Аня готовит реферат по истории и находит блог с профессиональным дизайном и уверенными формулировками. Она применяет CRAAP-тест: проверяет автора — его нет, дату — не указана, цель — продажа курса. Аня понимает, что источник недостоверен, и ищет другой.

### Пять критериев CRAAP




Каждый критерий CRAAP отвечает на свой вопрос. Вместе они дают полную картину надёжности источника.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Актуальность и релевантность: когда источник подходит

**Актуальность (Currency)** оценивает своевременность информации: когда материал был опубликован, обновлялся ли он, требуют ли темы исследования свежих данных. В быстро развивающихся областях — технологиях, медицине, праве — информация даже двух-трёхлетней давности может быть устаревшей. Однако в гуманитарных науках и истории актуальность менее критична: классические работы остаются ценными независимо от возраста.

**Релевантность (Relevance)** определяет, насколько информация соответствует теме исследования. Даже безупречно достоверный источник может оказаться неподходящим, если он адресован не той аудитории. Рецензируемая статья об изменении климата в Антарктиде будет неуместна для исследования городских островов тепла.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Авторитетность и точность: кто стоит за словами

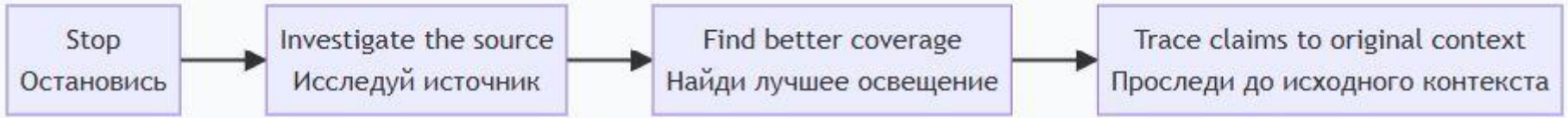
**Авторитетность (Authority)** оценивает квалификацию автора, издателя или организации. Выясните credentials автора, его институциональную принадлежность, образование и опыт. Уважаемые авторы часто цитируются в других источниках, поэтому обращайте внимание на имена, встречающиеся в разных публикациях.

В 2026 году проверяйте, существует ли автор вообще: ИИ может фабриковать имена. Ищите автора на ORCID, Google Scholar или институциональных страницах. Остерегайтесь «фабрик авторов», которые приписывают авторство реальным исследователям без их ведома.

**Точность (Accuracy)** означает, что информация фактически верна и подкреплена доказательствами. Ключевые вопросы: откуда взята информация, поддерживается ли она свидетельствами, прошла ли рецензирование, можно ли проверить данные в других источниках. Рекомендуется перекрёстная проверка: ищите подтверждение ключевых утверждений как минимум в двух других источниках.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## SIFT: четыре шага вместо чек-листа




Метод SIFT, предложенный Майком Коулфилдом, — эвристическая альтернатива развёрнутым чек-листам.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## SIFT: четыре шага вместо чек-листа

| |
|---|
| Остановитесь и не читайте дальше, пока не оцените источник |
| Откройте новую вкладку и поищите название сайта или автора |
| Посмотрите, что пишут о нем независимые источники |
| Если нашли лучшее освещение темы, переходите к нему |
| Проследите ключевые утверждения до исходного контекста |

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Почему SIFT работает: горизонтальное чтение

В 2004 году Марк Меола в статье «Chucking the Checklist» раскритиковал чек-листный подход к оценке информации. Позже Майк Коулфилд, критиковавший некоторые применения CRAAP-теста, предложил эвристический метод **SIFT** — четыре быстрых шага вместо развёрнутого чек-листа.

Первый шаг — **Stop** (остановись): не читайте дальше, пока не оцените источник. Второй — **Investigate the source** (исследуй источник): узнайте, кто автор и какова его репутация. Третий — **Find better coverage** (найди лучшее освещение): поищите, что говорят другие источники о той же теме. Четвёртый — **Trace claims, quotes, and media to the original context** (проследи утверждения, цитаты и медиа до исходного контекста): проверьте, не искажён ли смысл.

Сравнительное исследование 2019 года показало, почему SIFT работает: профессиональные фактчекеры за секунды открывали 5–6 вкладок, проверяя, что другие источники говорят о сайте (горизонтальное чтение), тогда как историки тратили минуты на чтение одного сайта (вертикальное чтение).

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Ловушки мозга: почему мы верим недостоверному

Наш мозг эволюционно не приспособлен к оценке цифровых источников, и два когнитивных искажения мешают особенно сильно. Эффект привязки (anchoring effect) заключается в том, что первая встреченная информация становится «якорем», и все последующие данные оцениваются относительно него. Подтверждающее искажение (confirmation bias) — это склонность верить информации, которая подтверждает уже имеющиеся убеждения, и скептически относиться к противоречащей.

Другие распространённые искажения включают эффект доступности — предпочтение информации, которая легко вспоминается, даже если она нерепрезентативна, и эффект группы — тенденцию принимать точку зрения большинства без критической оценки.

Логические ошибки также искажают оценку: ошибка ложной причины (post hoc, ergo propter hoc), чрезмерное обобщение на основе малого количества данных, аргумент к авторитету вместо фактических доказательств.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Практический алгоритм: от 30 секунд до 15 минут

| | |
|---|---|
| 1 | **Первичный отбор (30 секунд)**<br>Определите автора, дату публикации, издателя и цель, прежде чем углубляться в чтение. Если источник анонимный или дата отсутствует — это уже повод насторожиться. |
| 2 | **Исследование автора (2–3 минуты)**<br>Найдите информацию об авторе в интернете, проверьте его институциональную принадлежность, другие публикации и экспертность в данной теме. У академических авторов должны быть соответствующие степени или профессиональный опыт. |
| 3 | **Оценка публикации (2–3 минуты)**<br>К авторитетным издателям относятся университетские издательства, признанные академические издательства (Springer, Elsevier, SAGE), государственные ведомства и уважаемые аналитические центры. Будьте осторожны с хищническими издателями и неизвестными веб-сайтами. |

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Практический алгоритм: от 30 секунд до 15 минут

| | |
|---|---|
| 4 | **Анализ содержания (10–15 минут)**<br>Проверьте ссылки на источники, методологию, размер выборки, источники финансирования и потенциальную предвзятость. Убедитесь, что ссылки ведут на контент, который действительно поддерживает заявленные утверждения. |
| 5 | **Контекст цитирования (5 минут)**<br>Перепроверьте ключевые утверждения, находя подтверждение в других достоверных источниках. Правило триангуляции: если факт есть только на одном сайте — вероятно, это выдумка или ошибка. |

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Красные флаги недостоверных источников

**Нет автора или редакции** — анонимные статьи значительно чаще содержат ошибки

**Эмоциональные заголовки** — «ШОК!», «Вся правда о...» — маркеры кликбейта и манипуляции

**Нет ссылок на источники** — утверждения без ссылок — мнение, не факт

**Крайние формулировки** — эмоциональный, провокационный язык указывает на предвзятость

**Категоричные утверждения** — «всегда», «никогда», «все знают» — чрезмерное упрощение

**Противоречие консенсусу** — конфликт с устоявшимися исследованиями без веских доказательств

**Единственный источник** — факт есть только на одном сайте — вероятно, выдумка; проверьте минимум 3 независимых источника

Если вы заметили хотя бы два-три таких флага – источник требует особо пристальной проверки или отклонения.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Разбор: как применить все инструменты сразу

**КОНТЕКСТ.** Вы готовите исследование о городских островах тепла в мегаполисе. В поисковой выдаче первым идёт сайт с доменом .org, заголовком «ШОК! Города убивают климат!» и без указания автора. В тексте — эмоциональные утверждения без ссылок на источники.

**АНАЛИЗ.** Применяем CRAAP-тест: актуальность — дата публикации не указана; релевантность — тема близка, но уровень материала неясен; авторитетность — автор отсутствует; точность — нет ссылок и доказательств; цель — убедить, а не информировать. Затем применяем метод SIFT: исследуем источник — сайт принадлежит неизвестной организации; ищем лучшее освещение — находим рецензируемую статью по теме с методологией и ссылками.

**РЕЗУЛЬТАТ.** Источник отклоняется. Рецензируемая статья даёт проверяемые данные, которые можно перепроверить. Этот пример показывает: даже тематически релевантный источник без автора и доказательств — красный флаг, и системный подход помогает отсеять его за пару минут.

# ОЦЕНКА ДОСТОВЕРНОСТИ ИСТОЧНИКОВ

## Ключевые выводы

| |
|---|
| CRAAP-тест — пять критериев (актуальность, релевантность, авторитетность, точность, цель), разработан Сарой Блейксли в 2004 году. |
| Актуальность критична в быстро развивающихся областях, но менее важна в гуманитарных науках; проверяйте отзывы статей через Retraction Watch. |
| Домен не гарантирует достоверность — проверяйте автора через ORCID и Google Scholar. |
| Метод SIFT предлагает четыре быстрых шага: остановись, исследуй источник, найди лучшее освещение, проследи до исходного контекста. |
| Когнитивные искажения — эффект привязки и подтверждающее искажение — мешают объективной оценке. |
| Красные флаги: отсутствие автора, эмоциональные заголовки, отсутствие ссылок, категоричные утверждения. |

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ



# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Почему обычный поиск не справляется с соцсетями

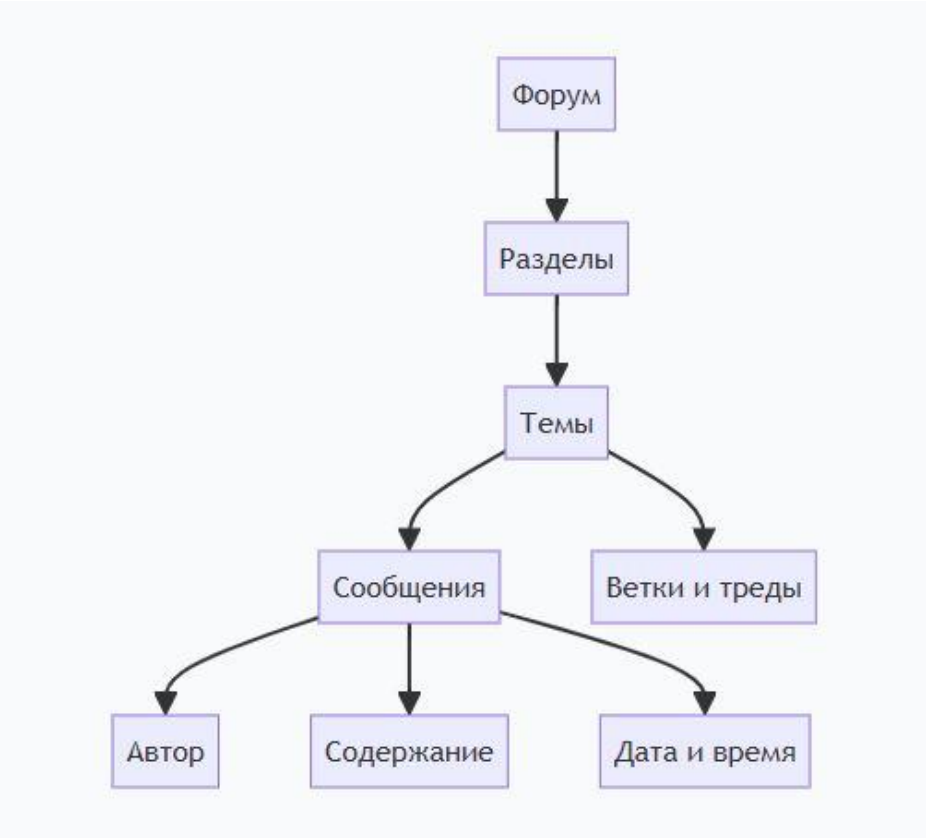
Традиционные поисковые системы, такие как Google, не индексируют большинство социального контента. Разговор, происходящий прямо сейчас на Reddit, набирающая популярность ветка на Twitter/X или запрос рекомендаций на LinkedIn — всё это не отображается в результатах Google до тех пор, пока не пройдут часы или дни, если вообще отобразится.

Причина в том, что классические поисковики созданы для статичных страниц, тогда как социальные сети — это быстрые, живые обсуждения, которые напрямую формируют покупательские решения.

Решение — социальные поисковые движки (social media search engines). Это инструменты, которые индексируют и ищут контент через API платформ и реальные потоки данных, возвращая результаты в едином интерфейсе.

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Анатомия форума: от разделов до сообщений



# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Древовидные и табличные форумы: две стратегии поиска

Веб-форум — это интернет-сервис для общения между пользователями на одну или несколько тем, и он является одной из форм социальных сетей. Самая распространённая иерархия: Разделы → Темы → сообщения (посты). Сообщения обычно имеют вид «автор — тема — содержание — дата/время» и образуют «ветки» или «темы».

Существуют два основных типа форумов. В древовидных форумах сообщения образуют строгую иерархию — можно чётко проследить, кто кому отвечает. В табличных (ленточных) форумах сообщения публикуются друг под другом по мере поступления.

Понимание этой структуры позволяет выбирать правильную стратегию: в табличных форумах достаточно просмотреть ленту, в древовидных — целенаправленно искать по веткам. Это как разница между чтением газеты и изучением генеалогического древа — подходы принципиально разные.

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## OSINT-инструменты для поиска пользователей

**Namechk**

Проверка имени пользователя на множестве платформ, поиск персон в интернете

**Sherlock**

Open-source CLI-инструмент на Python, ищет имя пользователя в популярных соцсетях

**Accountanalysis**

Отчёт по Twitter-аккаунту: подписчики, частота твитов, время активности, хештеги

**Maltego**

Графы связей между IP, доменами, email и пользователями соцсетей

**UserSearch**

Профессиональная OSINT-платформа: 17 типов поиска — имя, email, телефон, фото, криптовалюта

Пять инструментов для разных задач: от быстрой проверки имени до построения графа связей.

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Как работают OSINT-инструменты

OSINT-инструменты — это специализированные сервисы для сбора данных из открытых источников. **Namechk** позволяет проверить наличие имени пользователя на различных онлайн-платформах — достаточно ввести имя, и сервис прочешет множество платформ в поисках соответствующих профилей. **Sherlock** — инструмент командной строки с открытым исходным кодом на Python, который, подобно Namechk, ищет определённое имя пользователя в наиболее популярных социальных сетях и выдаёт ссылки на профили.

**Maltego** — один из самых мощных инструментов OSINT. Он использует графики для представления информации и позволяет выявлять закономерности и скрытые связи между сущностями: IP-адресами, доменными именами, электронной почтой, пользователями соцсетей.

Выбор инструмента зависит от задачи: нужно быстро проверить имя — Namechk или Sherlock; понять активность аккаунта — Accountanalysis; выявить связи — Maltego; провести профессиональное расследование — UserSearch.

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Кросс-платформенные социальные поисковые движки

Социальные поисковые движки — это инструменты, которые индексируют и ищут контент через API платформ и реальные потоки данных, возвращая результаты в едином интерфейсе. Они отличаются от традиционных поисковиков тем, что фокусируются на пользовательском контенте: постах, твитах, фотографиях, видео, профилях.

**Hootsuite** позволяет искать в реальном времени на 30+ социальных сетях и 150+ миллионах веб-сайтов, включая форумы, подкасты, новостные сайты, печатные и вещательные СМИ, в 239 странах и 186 языках. Встроенный ИИ и умный конструктор запросов помогают получать точные результаты без сложных булевых операторов.

**Talkwalker Free Social Search** анализирует разговоры в реальном времени на 186 языках по всем основным социальным сетям и позволяет фильтровать результаты по местоположению, языку, типу медиа, тональности.

**Brandwatch** — корпоративный инструмент, работающий со 100+ миллионами источников и 1,4 триллионами постов, включая социальные сети, блоги, форумы и новостные сайты.

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Методология поиска тем на форумах для контента

| | |
|---|---|
| 1 | **Выберите популярный форум по вашей тематике**<br>Например, автомобильной или женской. Если подходящего форума не нашлось, используйте сайты вопросов-ответов, такие как Ответы@Mail.ru, и ищите по ключевым словам. |
| 2 | **Найдите раздел по нужной теме**<br>Изучите структуру форума: разделы, подразделы, сверхразделы. Это поможет понять, где сосредоточена нужная информация. |
| 3 | **Просмотрите новые темы, популярные темы или темы с недавними ответами**<br>Это покажет, что сейчас актуально для аудитории и какие вопросы волнуют пользователей. |

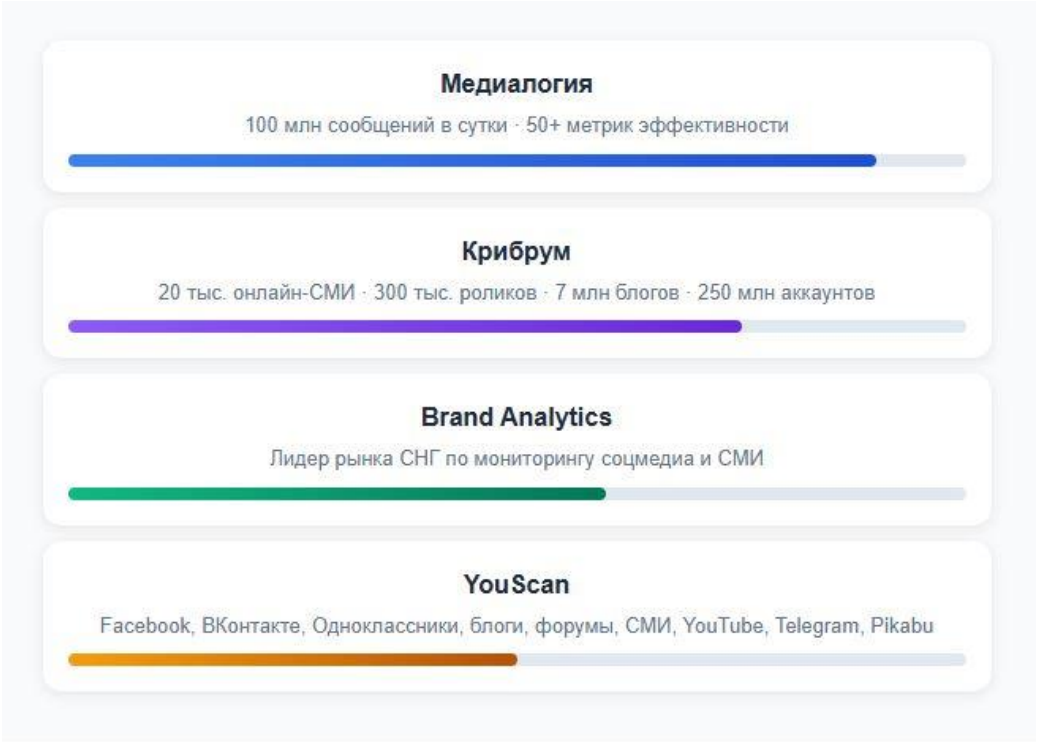
# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Методология поиска тем на форумах для контента

| | |
|---|---|
| 4 | **Используйте фильтры**<br>На всех форумах есть фильтры, позволяющие сортировать темы по различным параметрам — дате, количеству ответов, просмотрам. |
| 5 | **Обратите внимание на названия тем**<br>Они часто звучат как готовые заголовки для статей или постов. Уже сами названия тем подсказывают множество идей. |
| 6 | **Читайте ответы пользователей и выписывайте понравившиеся темы**<br>Даже если они звучат «кривенько» — потом их можно переделать. Из ответов пользователей можно сделать несколько смежных тем. Пример из практики: вопрос пользовательницы на форуме может стать основой для серии постов: «Помогает ли пилинг убрать последствия прыщей?», «Стоит ли маскировать подростковые проблемы с кожей ВВ-кремом?», «ВВ-крем и подростковая кожа», «Мешает ли косметика лечению прыщей?». |

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Сервисы мониторинга: масштаб данных




Масштаб данных, которые обрабатывают ведущие сервисы мониторинга.

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Сервисы мониторинга социальных сетей

Сервисы мониторинга социальных сетей позволяют отслеживать упоминания бренда, продуктов или услуг в социальных медиа, анализировать тональность этих упоминаний и реагировать на них. Такие системы, как LiveDune, SCAN, Brand Analytics, Медиалогия, SocialVisor, ExLibris, DataLocator, собирают данные из соцсетей, форумов, блогов и СМИ.

**Brand Analytics** — лидирующая на рынке СНГ система мониторинга и анализа социальных медиа и СМИ. Она собирает обратную связь из соцмедиа и помогает понять потребности аудитории. **Медиалогия** обрабатывает 100 миллионов сообщений в сутки и предоставляет более 50 метрик эффективности для мгновенной оценки тональности упоминаний.

**Крибрум** анализирует данные из 20 тысяч онлайн-СМИ, 300 тысяч роликов, 7 миллионов блогов и 250 миллионов аккаунтов в соцсетях. **YouScan** отслеживает упоминания в Facebook, ВКонтакте, Одноклассниках, блогах, форумах, СМИ, на сайтах отзывов, YouTube, в Telegram-каналах и на Pikabu.

# ПОИСК В СОЦИАЛЬНЫХ СЕТЯХ И НА ФОРУМАХ

## Ключевые выводы

| |
|---|
| Google не индексирует большинство социального контента — для поиска в соцсетях нужны специализированные инструменты, работающие через API платформ. |
| Структура форума (Разделы → Темы → сообщения) определяет стратегию поиска: в табличных форумах просматривают ленту, в древовидных — ищут по веткам. |
| OSINT-инструменты (Namechk, Sherlock, Accountanalysis, Maltego, UserSearch) решают разные задачи: от проверки имени до построения графа связей. |
| Кросс-платформенные движки (Hootsuite, Talkwalker, Octolens, Brandwatch) ищут по всем соцсетям в реальном времени в едином интерфейсе. |
| Сервисы мониторинга (Brand Analytics, Медиалогия, Крибрум, YouScan) отслеживают упоминания бренда и анализируют тональность. |
| Методология поиска тем на форумах: выбор форума, разделы, фильтры, названия тем и ответы пользователей — источник идей для контента. |

# ПРОФЕССИОНАЛЬНЫЙ ПОИСК: ПРАВО, МЕДИЦИНА, ТЕХНИКА



КонсультантПлюс

надежная правовая поддержка

# ПРОФЕССИОНАЛЬНЫЙ ПОИСК: ПРАВО, МЕДИЦИНА, ТЕХНИКА

## Почему профессиональный поиск — это другой мир

Профессиональный поиск отличается от обычного не набором операторов, а самой природой данных. Когда вы ищете «как составить договор» в Google, вы получаете тысячи статей и блогов. Но юристу нужен не текст о договоре, а действующий нормативный акт, форма документа или судебная практика — и всё это хранится в специализированных правовых базах, таких как КонсультантПлюс.

То же самое в медицине и технике. Проверка квалификации врача или инженера — это не поиск информации, а верификация статуса. Лицензионные базы данных хранят записи о профессиональных лицензиях, выданных государственными органами. Эти системы имеют собственную структуру, фильтры и логику, отличную от общих поисковиков.

Ключевой принцип: в профессиональном поиске вы ищете не «что-то о теме», а конкретный тип документа или конкретный статус. Именно поэтому понимание внутренней структуры базы становится важнее, чем умение формулировать запрос.

# ПРОФЕССИОНАЛЬНЫЙ ПОИСК: ПРАВО, МЕДИЦИНА, ТЕХНИКА

## КонсультантПлюс: база, где тип документа важнее текста

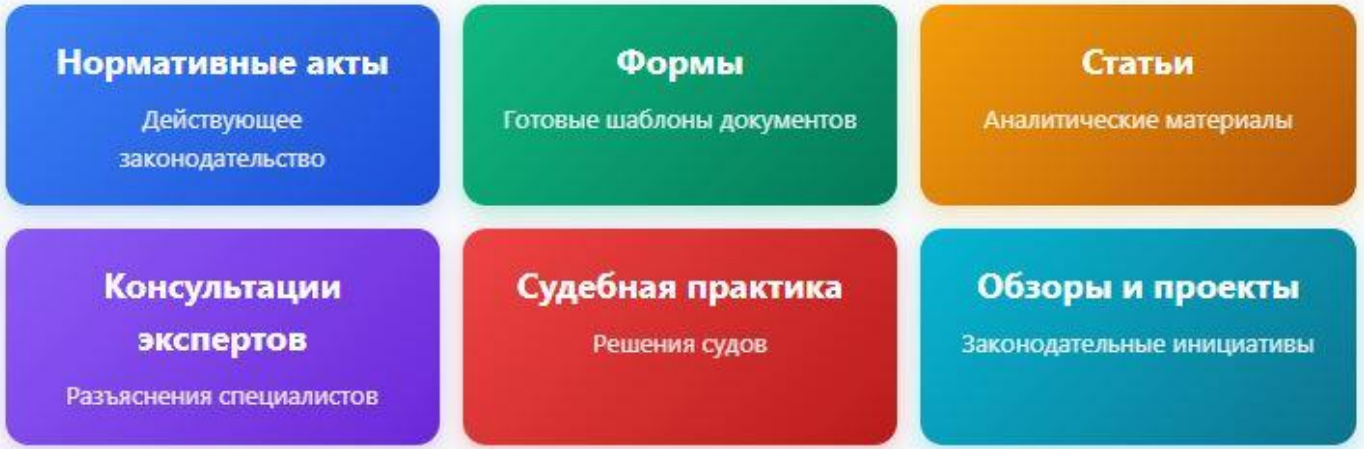
КонсультантПлюс — это не просто поисковая система по законодательству, а структурированная база, где каждый документ относится к определённому типу. Здесь есть нормативно-правовые акты, формы документов, статьи, консультации экспертов и судебная практика. Для профессионала это означает, что можно искать не «что-то про договоры», а сразу судебную практику по договорам или готовую форму договора.

Выбор типа документа зависит от задачи. Нужно подготовить договор — ищите формы. Нужно понять, как закон применяется на практике — обращайтесь к судебной практике. Нужны разъяснения по сложному вопросу — ищите консультации экспертов. Система даже формирует подборки документов по конкретным темам — например, по запросу «Профстандарты в здравоохранении» собираются все связанные материалы.

Помимо основных типов, база содержит обзоры федерального и регионального законодательства, проекты правовых актов, справочную информацию, календари и формы документов.

# ПРОФЕССИОНАЛЬНЫЙ ПОИСК: ПРАВО, МЕДИЦИНА, ТЕХНИКА

## Типы документов в КонсультантПлюс




Каждый тип документа в КонсультантПлюс решает свою задачу — от подготовки договора до отслеживания законодательных инициатив.

# ПРОФЕССИОНАЛЬНЫЙ ПОИСК: ПРАВО, МЕДИЦИНА, ТЕХНИКА

## Лицензия и сертификация: два разных статуса

В профессиональном поиске важно различать два статуса, которые часто путают: лицензирование и сертификацию. Лицензия — это разрешение на деятельность, выданное государственным органом. Без неё практиковать нельзя. Сертификация — это добровольное подтверждение углублённой квалификации в узкой области.

Почему это различие критично? Представьте специалиста с впечатляющей сертификацией, но без лицензии. Формально он не имеет права заниматься практикой. И наоборот: наличие лицензии — это базовое требование, а сертификация — дополнительный показатель экспертизы.

Для профессионального поиска это означает, что оба статуса нужно проверять отдельно. База данных может показать наличие лицензии, но не отразить сертификацию, или наоборот. Понимание разницы помогает избежать ошибок при оценке квалификации.

# ПРОФЕССИОНАЛЬНЫЙ ПОИСК: ПРАВО, МЕДИЦИНА, ТЕХНИКА

## Профессиональные домены

### Здравоохранение

Врачи, медсёстры, клинические психологи, стоматологи

### Инженерия и архитектура

Инженеры-строители, ландшафтные архитекторы, геодезисты, домашние инспекторы

### Недвижимость

Специалисты по операциям с недвижимостью

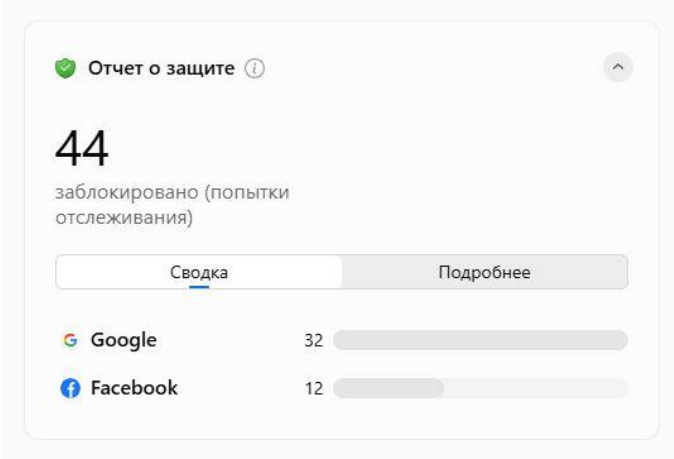
### Бухгалтерия и бизнес

Бухгалтеры и консультанты по бизнес-услугам

Лицензионные базы организуют информацию по профессиональным доменам — от здравоохранения до бухгалтерии.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ



# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Ваш поисковый запрос — это не секрет

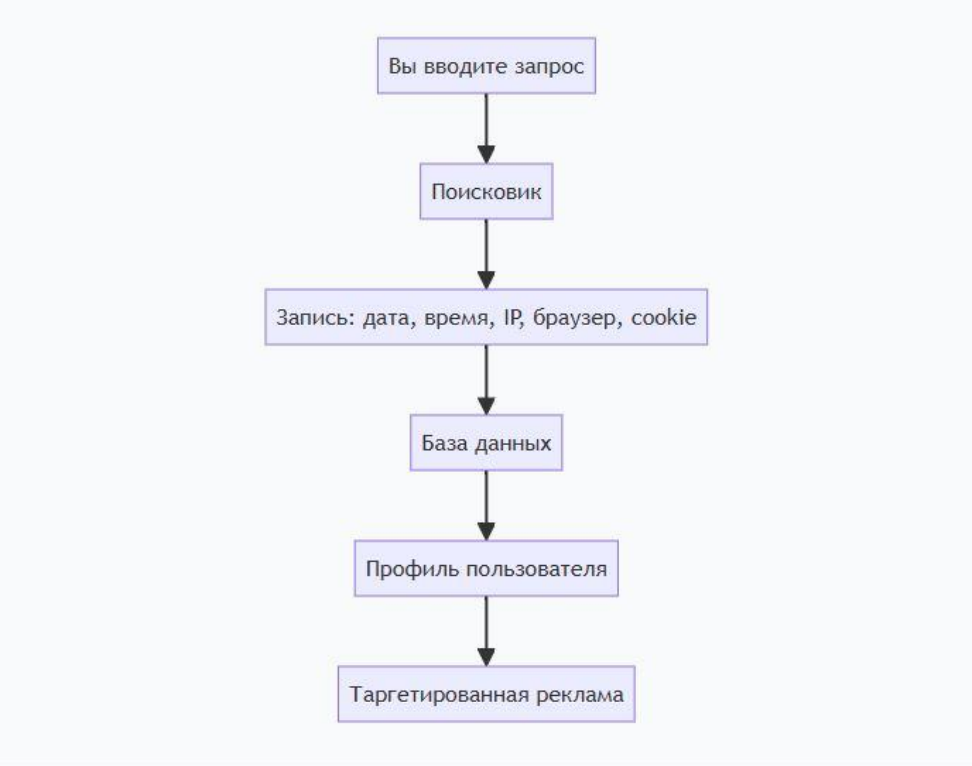
Каждый раз, когда вы вводите запрос в Google, вы оставляете цифровой след: дату, время, IP-адрес, тип браузера, cookie-файлы. Поисковик записывает всё это в базу данных.

Традиционные поисковые системы, такие как Google и Yahoo, собирают значительные объёмы данных. Google регистрирует все поисковые термины в базе данных вместе с датой и временем поиска, типом браузера и операционной системы, IP-адресом пользователя, cookie-файлом Google и URL-адресом, показывающим поисковую систему и поисковый запрос. Это не просто техническая необходимость — эти данные используются для построения поведенческих профилей.

Зачем это нужно? Поисковики утверждают, что собирают данные для улучшения релевантности результатов. Однако они могут злоупотреблять конфиденциальностью, продавая информацию рекламодателям.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Путь ваших данных




Каждый запрос превращается в запись, которая затем используется для профилирования и рекламы.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Политики конфиденциальности: почему вы не читаете их

Политики конфиденциальности структурированы по модели «уведомление и согласие»: сайт показывает вам политику, и вы кликаете «согласен». Однако это решение не является свободным, поскольку издержки отказа могут быть очень высокими. Если вы откажетесь, вы не сможете пользоваться сервисом.

Кроме того, политики часто сложны для понимания, даже если вы решите их прочитать. Многие люди никогда не читают политики конфиденциальности и поэтому не осознают, сколько их частной информации — паролей и сохранённых файлов — собирается через cookie-файлы и может регистрироваться и храниться поисковой системой.

Приватные поисковые системы, такие как DuckDuckGo, заявляют в своих политиках, что собирают гораздо меньше данных, чем Google или Yahoo, а возможно, и вовсе ничего. Это делает их привлекательной альтернативой для тех, кто ценит конфиденциальность.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Традиционные vs приватные поисковики

| ТИП | СБОР ДАННЫХ | ПРИМЕРЫ |
|---|---|---|
| **Традиционные** | IP-адрес, cookie, история запросов, профилирование | Google, Yahoo |
| **Приватные** | Минимум данных, анонимизация запросов | DuckDuckGo, Startpage, Searx |

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## DuckDuckGo и Startpage: как они защищают вас

DuckDuckGo — наиболее используемая частная поисковая система в открытом доступе и самый простой вариант для перехода с Google. Результаты поиска поступают в основном от Bing, при этом все идентифицирующие данные удаляются перед запросом — Bing видит IP-адрес DuckDuckGo, а не пользователя. DuckDuckGo не собирает и не передаёт личную информацию пользователей, такую как IP-адреса или cookie-файлы, не имеет спама и анонимизирует поисковые запросы от выбранного пользователем веб-сайта, используя шифрование.



# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## DuckDuckGo и Startpage: как они защищают вас

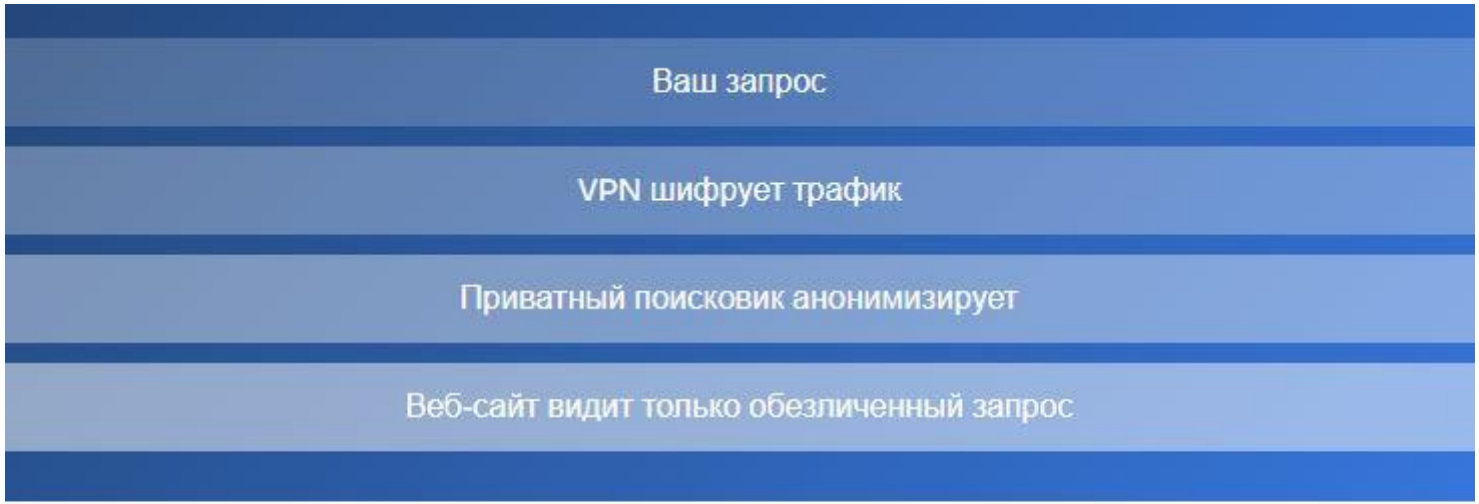
Одна из уникальных функций DuckDuckGo — библиотека сочетаний клавиш bang (более 13 000 сайтов). Например, !w переходит на Википедию, !yt — на YouTube, !a — на Amazon. Это позволяет искать напрямую на других сайтах, не покидая поисковик.

Startpage — анонимный поисковик, использующий результаты Google без трекинга. В функции поисковика входят прокси, генератор URL и поддержка HTTPS. Генератор URL — уникальная функция, позволяющая запоминать настройки без cookie-файлов.

Также стоит упомянуть Searx — метапоисковую систему с открытым кодом, которая собирает результаты популярных поисковых систем и объединяет их. Searx удаляет все идентифицирующие данные из запроса, так что Google, Yahoo и другие получают поисковую фразу в виде анонимного запроса. Система не хранит никаких данных о поиске и никогда не передаёт ничего третьим лицам.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Слои защиты




Каждый слой добавляет дополнительную защиту, но ни один не делает вас полностью невидимым.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Проверка приватности поисковика

Проверка приватности поисковика — это не один тест, а несколько независимых уровней. Разберу по порядку.

Первый уровень — **политика конфиденциальности**. Читай не вводный абзац, а разделы про то, какие данные собираются, как долго хранятся и кому передаются. Ключевые маркеры: сохраняется ли IP-адрес, привязывается ли история к аккаунту, есть ли срок хранения в днях. Расплывчатые формулировки вроде «можем использовать данные для улучшения сервиса» — сигнал, что заявленная приватность ограничена.

Второй уровень — **технический анализ трафика**. Открой инструменты разработчика в браузере и посмотри, какие запросы уходят при загрузке страницы и при вводе запроса. Если поисковик обещает не логировать, но отправляет данные на сторонние домены аналитики — обещание не соответствует поведению.

Третий уровень — **исходный код**. У поисковиков с открытым кодом можно проверить логику обработки запросов напрямую, а не верить описанию. У закрытых сервисов такой возможности нет, и это само по себе ограничение.

Четвёртый уровень — **независимые аудиты**. Если сервис проходил внешнюю проверку и опубликовал отчёт — это весомее собственных заявлений. Отсутствие аудита не значит, что тебя обманывают, но и подтверждением приватности не является.

И честное ограничение: полностью доказать отсутствие сбора данных со стороны пользователя невозможно. Ты видишь только внешнее поведение и заявления, но не внутреннюю инфраструктуру. Поэтому разумнее оценивать не «точно ли ничего не собирают», а насколько сервис снижает объём собираемых данных по сравнению с альтернативами.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Ограничения приватных поисковиков и дополнительные меры

Приватные поисковые системы уменьшают профилирование, ограничивая личные идентификаторы, сохранение истории поиска или отслеживание рекламы, но они не делают вас невидимым: открываемые веб-сайты, интернет-провайдер, расширения браузера и поисковая система по-прежнему могут отслеживать часть пути.

Режимы инкогнито в Chrome или Firefox не обеспечивают тот уровень конфиденциальности, который ожидают пользователи. Они лишь не сохраняют историю на вашем устройстве, но ваш IP-адрес и данные всё равно видны.

Для дополнительной защиты используются VPN-сервисы, которые шифруют весь трафик, чтобы третьи лица — интернет-провайдер, хакеры и государственные учреждения — не могли видеть ваши действия. Также помогают tracking-блокеры, такие как uBlock Origin или Privacy Badger, и безопасные браузеры, включая Tor или Brave.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Как работает Results About You

**Results about you** («Результаты о вас») — это инструмент от Google, который помогает контролировать свою цифровую репутацию: находить в поисковой выдаче личную информацию и запрашивать её удаление.

С его помощью можно найти, например, такие данные: номер телефона, домашний или рабочий адрес, email, сведения из паспорта или водительских прав, номер страховки.

Инструмент Results About You позволяет запросить удаление информации одной кнопкой. После настройки Google отправляет электронные письма при обнаружении имени, электронной почты, номера телефона или адреса.

Важно понимать: инструмент не стирает информацию на других веб-страницах, а только удаляет ссылки из Google Search, однако это создаёт препятствия и делает вас менее лёгкой целью.

# КОНФИДЕНЦИАЛЬНОСТЬ И БЕЗОПАСНОСТЬ

## Ключевые выводы

| |
|---|
| Традиционные поисковики собирают IP-адреса, cookie, историю запросов и строят поведенческие профили. |
| Политики конфиденциальности по модели «уведомление и согласие» часто не оставляют реального выбора. |
| Приватные поисковики (DuckDuckGo, Startpage, Searx) минимизируют сбор данных и анонимизируют запросы. |
| Приватные поисковики не делают вас невидимым: сайты, провайдер и расширения могут отслеживать часть пути. |
| Инструмент Google Results About You помогает удалять личную информацию из результатов поиска, но не стирает её с других сайтов. |

# СОЗДАНИЕ ПЕРСОНАЛЬНОЙ СИСТЕМЫ ПОИСКА



# СОЗДАНИЕ ПЕРСОНАЛЬНОЙ СИСТЕМЫ ПОИСКА

## Дерево папок против тегов

Традиционная структура папок — это дерево: каждая папка может содержать только определённые файлы, если не считать копирования и ссылок. Поиск по названию файла требует строгой последовательности слов, а количество символов в названии ограничено.

Представьте: у вас есть папка с музыкой и подпапки для жанров. Но что, если вы захотите послушать французскую музыку независимо от жанра? Дерево не позволяет этого сделать без создания отдельных папок под язык и ссылок. Решение — тегирование: каждому файлу присваиваются теги с жанром, языком и типом контента. Комбинируя три тега (например, «французская», «рок»), вы получаете то, чего стандартными средствами достичь невозможно.

Тегирование позволяет группировать и сортировать информацию гораздо гибче, независимо от порядка слов в запросе. Это не просто удобство, а смена парадигмы: вместо жёсткой иерархии вы получаете многомерную группировку, где каждый файл может принадлежать сразу нескольким категориям.

# СОЗДАНИЕ ПЕРСОНАЛЬНОЙ СИСТЕМЫ ПОИСКА

## Сравнение поиска

| Поиск по названию файла | | Поиск по содержимому (индекс) |
|---|---|---|
| Важен | **Порядок слов** | Не важен |
| Нет | **Ищет по содержимому** | Да |
| Нет | **Требует индексации** | Да |
| Низкая | **Точность при неполном названии** | Высокая |

# СОЗДАНИЕ ПЕРСОНАЛЬНОЙ СИСТЕМЫ ПОИСКА

## Архитектура персональной системы поиска




Компоненты персональной системы поиска: от данных к визуализации.

# СОЗДАНИЕ ПЕРСОНАЛЬНОЙ СИСТЕМЫ ПОИСКА

## Индексация: как система узнаёт ваши документы

Мы уже знаем, как работают поисковые системы: ввод запроса, получение данных из индекса, формирование результатов. Индекс — это база данных со списком всех документов, которые система знает. Для персональной системы принцип тот же: необходимо создать индекс всех личных документов и файлов.

При создании частотного словаря и обратного индекса игнорируются стоп-слова — слова, которые не несут значимой информации, так как есть почти в каждом документе (предлоги, союзы, указательные местоимения). Обратный индекс — это массив слов со ссылками на документы, в которых эти слова упоминаются. Это позволяет при запросе не обходить весь индекс, а сразу получить готовые подгруппы документов.

# СОЗДАНИЕ ПЕРСОНАЛЬНОЙ СИСТЕМЫ ПОИСКА

## Выбор ИИ-инструмента под задачу

| ИНСТРУМЕНТ | СИЛЬНАЯ СТОРОНА | КОМУ ПОДХОДИТ |
|---|---|---|
| ChatGPT | Сложные вопросы, исследования, пояснения | Исследователи, аналитики, преподаватели |
| DeepSeek | Научные запросы, точные цитаты и DOI | Учёные, студенты, библиографы |
| Perplexity | Ответы на основе множества источников со ссылками | Широкий круг задач |
| Gemini | Обзоры, справочная и прикладная информация | Повседневные запросы |
| Claude | Вдумчивый анализ и структурирование информации | Аналитики, авторы |
| Grok | Анализ новостей и текущих трендов | Журналисты, наблюдатели за трендами |

# СОЗДАНИЕ ПЕРСОНАЛЬНОЙ СИСТЕМЫ ПОИСКА

## Стратегии поиска: план, а не случайность

Функционирование информационно-поисковых систем основывается на двух предположениях: нужные документы объединены наличием некоторого характеристического признака, и пользователь способен указать этот признак. На практике оба предположения редко выполняются, поэтому поиск — это последовательность шагов, качество которых имеет случайную природу.

Стратегия поиска — это общий план поведения пользователя для выражения и удовлетворения информационной потребности. Выбор стратегии — оптимизационная задача, которая на практике определяется искусством достижения компромисса между потребностями и возможностями средств.

Существуют различные классы стратегий: случайная (последующий вариант формулировки не связывается с предыдущими), стратегия расширения, стратегия сужения и смешанная (композиционная) стратегия. В персональной системе выбор стратегии становится особенно важным, поскольку пользователь сам управляет всеми параметрами.

# СОЗДАНИЕ ПЕРСОНАЛЬНОЙ СИСТЕМЫ ПОИСКА

## Ключевые выводы

| |
|---|
| Тегирование решает ограничения древовидной структуры папок, позволяя комбинировать признаки независимо от иерархии. |
| Google CSE — быстрый старт для создания поисковой системы без разработки собственных алгоритмов. |
| Собственное приложение: хранение в localstorage с синхронизацией через MySQL обеспечивает доступность с любого устройства. |
| Индексация в персональной системе строится на обратном индексе, стоп-словах и оценке TF×IDF. |
| ИИ-инструменты (ChatGPT, DeepSeek, Perplexity, Gemini, Claude, Grok) подбираются под тип задачи. |
| Стратегии поиска — случайная, расширение, сужение, смешанная — определяют план поведения пользователя. |

# ДОМАШНЕЕ ЗАДАНИЕ

Используя сервис Яндекс Формы ответить на вопросы:

- Открыть форму по ссылке: https://forms.yandex.ru/u/6ad35163852af39e93c7a286

  или

  https://clck.ru/3WZjLt
- Ответить на предлагаемые вопросы
- Если не набираете проходной балл, 75% верных ответов, проходите задание повторно.

Главный специалист отдела электронных ресурсов

Осипов Олег Геннадьевич

СОУНБ им. В.Г. Белинского

+7 343 304 60 15, доб. 320

z_inf@uraic.ru