

Поисковые системы в интернете:  
разновидность, особенности, преимущества  
и недостатки

Корнев А.Н

Исупов В.А

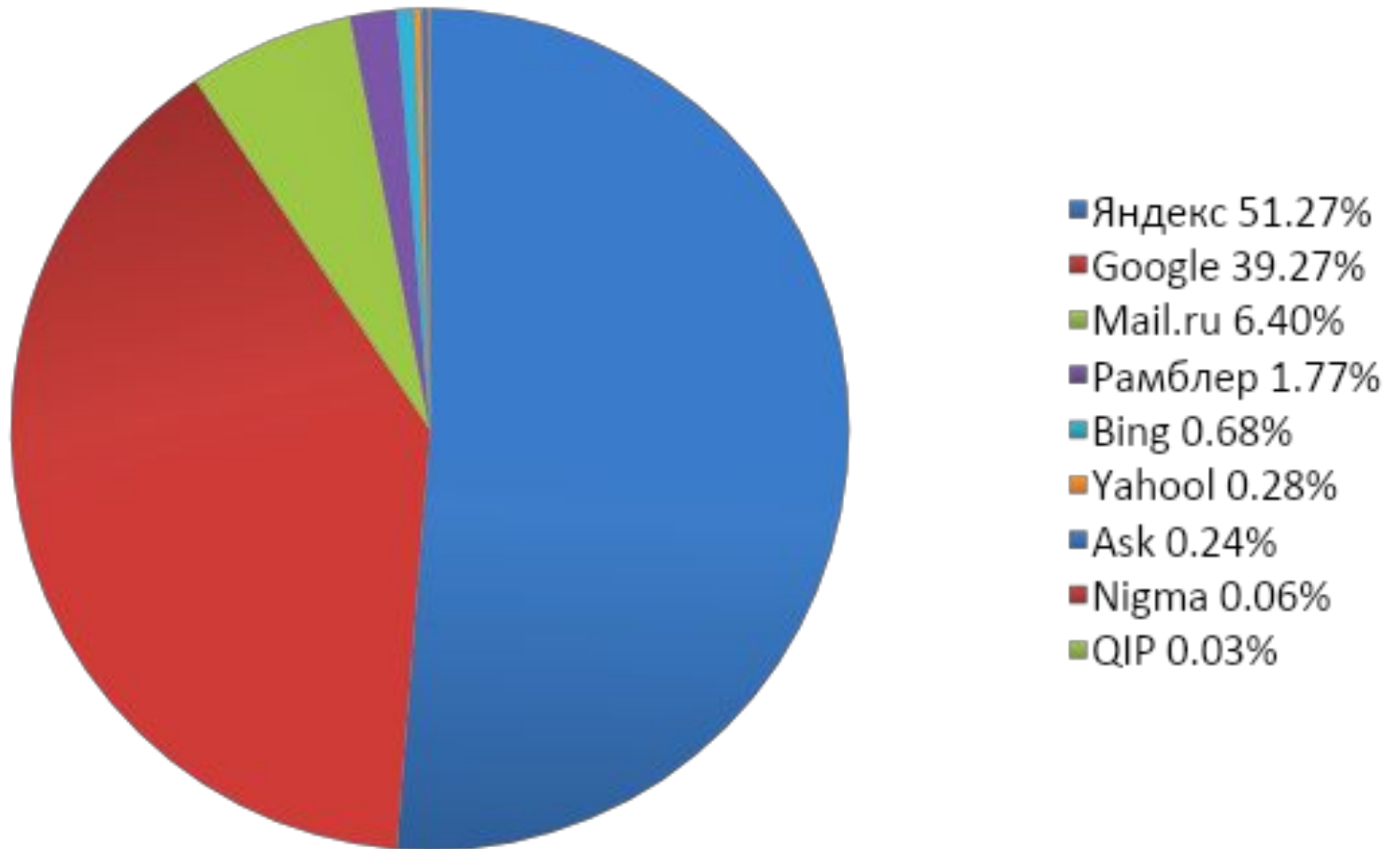
# Содержание

- [Рейтинг поисковых систем](#)
- [Компоненты поисковых систем](#)
- [Виды поисковых систем](#)

## Недостатки

- [Релевантность](#)
  - [Запросы на человеческом языке](#)
  - [Малоинформативные запросы](#)
  - [Синонимы](#)
  - [Спам](#)
- 
- [Информационные источники](#)

# Рейтинг поисковых систем



# Компоненты поисковых систем

А что же помогает поисковику, например, тому же Гугл индексировать ваш ресурс?

Агенты — это работники, выполняющий основную часть работы — индексируют и анализируют сайты.

Пауки (spider) – программа, которая может скачивать страницы веб-ресурса и собирать общую информацию о нем.

Кроулеры (crawler) – программа, отыскивающая все ссылки на страницах, переходя по которым разыскивает новые данные не знакомые поисковику.

Индексатор (indexer) – анализирует текст, заголовки, стиль и др.

Роботы - индексируют страницы вашего контента, а также изучают разнообразные ссылки.

Для того чтобы индексация происходила так, как нужно вам вы и создаете специальный документ «robots.txt». Он позволяет системе проверять только те страницы, которые вам нужно, и убирать то, что видеть не следует.



# Виды поисковых систем

Существуют несколько вариантов информационно — поисковых систем:

- Каталоги. Простое сравнение поиска – это книжная полка в библиотеке. Там все хранится в подкатегориях и категориях определенной тематики. Если вы попали в такой поисковик, то поверьте, информация, которую вы там найдете, будет более чем полезна и понятна для вашего восприятия. Догадались о каком распространенном сайте идет речь? Конечно же о википедии, которая собрала в себе целый справочник полезной информации.
- Поисковые указатели. Поиск в данных осуществляется за счет ключевых фраз. Это и удобно и неудобно одновременно. Думаю, меня поймут те люди, которые ищут, например, «Девушка показывает класс», чтобы найти как девушка показывает большой палец вверх, а в поиске вылезает что-то не очень приличное. Такой вид поиска характеризует большую часть поисковых систем.
- Рейтинговые системы. Определяют вашу популярность за счет количества посещений. Конечно, не самый лучший критерий, так как не всегда учитывается полезность и качество самого ресурса. Пример такой системы – это интернет ресурс alexa.com.
- Поисковые сервера подразделяются также на общие и специализированные. Общие поисковики сортируют информационные данные без всякого отбора по всем известным им веб-ресурсам. К ним относятся Яндекс, Рамблер, Гугл. Специализированные — осуществляют сортировку по используемому языку.

Также поисковые системы могут делиться на региональное и мировое распространение.

На сегодняшний день все поисковики постоянно совершенствуют свои алгоритмы по отбору качественных, релевантных ресурсов.



# Релевантность

Релевантность относительна. Это самая большая проблема. Вы можете искать “кофе” в Канаде и найти Tim Horton’s сайт, как наиболее релевантный. Это самая популярная сеть кофейных магазинов Канады, но для жителя Сиэтла самым релевантным будет Starbucks. Также вы можете ввести запрос “60-е года”, пытаться найти старую футбольную команду, а историк тем же запросом будет искать материал по истории Калифорнии. Вы можете искать сегодня “bones” (кости), чтобы узнать, где можно купить кость для собаки, а завтра тем же запросом будете искать новую серию сериала “Bones”.

Как поисковой системе разобраться в таких запросах? Только если вместо машины запрос будет читать человек...

До настоящего времени все, что имеют поисковые системы для решения этой проблемы - это память cookies и персональный поиск. Чем лучше поисковая система знает пользователя, тем выше шансы дать ему релевантный результат. Недостаток этого метода: обеспокоенность пользователей по поводу сохранности личных данных.

2. Запросы на человеческом языке. Движок базы данных MySQL может выдать все релевантные записи, выполняя команду `'select first,last from employee where last="Smith";'`. Это форматный синтаксис, исключая двусмысленность. Поисковая система, с другой стороны, получает строку `'who has smith as last name in chicago'` (у кого в Чикаго фамилия Smith) или `'smith last name chicago'`. Естественный язык запросов - наш язык, из слов люди составляют осмысленные выражения. Есть много способов сказать об одном и том же, есть разные виды личностей и т. д. Перед тем как искать что-либо в индексе, поисковая система должна разбить запрос на части и привести его в более форматный вид.



# Запросы на человеческом языке

Запросы на человеческом языке. Движок базы данных MySQL может выдать все релевантные записи, выполняя команду `'select first,last from employee where last="Smith";'`. Это форматный синтаксис, исключающий двусмысленность. Поисковая система, с другой стороны, получает строку `'who has smith as last name in chicago'` (у кого в Чикаго фамилия Smith) или `'smith last name chicago'`. Естественный язык запросов - наш язык, из слов люди составляют осмысленные выражения. Есть много способов сказать об одном и том же, есть разные виды личностей и т. д. Перед тем как искать что-либо в индексе, поисковая система должна разбить запрос на части и перевести его в более форматный вид.



# Малоинформативные запросы

Многие пользователи не могут толково объяснить, что они хотят в реальном мире, и еще хуже когда пытаются объяснить это поисковой системе. Они вводят 'пылесос', и поэтому не могут найти сервис для уборки дома. К тому же они делают ошибки в словах, создавая более "интересную" проблему для поисковой системы.





# Синонимы

Еще одна проблема. Есть слова с одинаковым значением, например “машина” и “автомобиль”. Когда вы ищете что-нибудь, вы хотите увидеть не только страницы, на которых упоминается ключевое слово, но и страницы с синонимами, если они будут релевантны вашему запросу. Скажем, вы ввели запрос ‘monkey’ (макака). Вы также возможно захотите увидеть и страницы, содержащие слова ‘chimpanzee’ или ‘ape’ (шимпанзе и горилла). Однако если вы будете подходить к запросу серьезно, возможно вы не захотите видеть страницы о шимпанзе, потому что хоть шимпанзе и приматы, но они не макаки. Такие проблемы редко приходят в голову пользователям, но для поисковых систем они довольно актуальны.



# Спам

Поисковые системы определяют релевантные страницы по “сигналам качества” либо параметрам, которые можно просчитать исходя из содержания страниц (ключевые слова, наличие в тегах title, b, strong, i и так далее - прим. egorych). Вызов для поисковых систем в том, что как только Blackhat Seo’s узнают об этих “сигналах”, они начинают их использовать (в своих корыстных целях : ) - прим. egorych). На мой взгляд со временем становится все труднее создавать эти фальшивые сигналы, но это никогда не станет невозможным. Для людей распознать спам просто, но для компьютера это намного сложнее.



# Информационные источники

- <https://seo-praktika.com/nedostatki-poiskovyh-sistem.html>
- <https://kner.ru/tech/list-of-search-engines.html>
- [https://ru.wikipedia.org/wiki/Поисковая\\_система](https://ru.wikipedia.org/wiki/Поисковая_система)



Сюда фотку “Спасибо за  
внимание!”