**2 слайд:**

Перед тем, как приступить к сравнительному анализу поисковых машин интернета, дадим определение понятию «поисковая машина».

Поисковая машина (поиско́вый движо́к) — комплекс программ, предназначенный для поиска информации. Обычно является частью поисковой системы и рассматриваются в ее контексте.

**3 слайд:**

К этому комплексу программ существуют определенные требования к функциональности:

* Поиск ссылок на страницы и другие документы сайтов.

Он может происходить в двух режимах:

* + Автоматический, когда поисковая машина ищет ссылки со страниц сайтов;
  + Ручной режим, когда пользователи сами добавляют в базу данных поисковой машины ссылки на страницы своих сайтов.
* Индексация документов сайтов:
  + Под этим понимается извлечение из документов информации, важной для поиска, преобразование этой информации в формат, удобный для поисковой машины и сохранение этой информации в базу данных поисковой машины.
* Поиск по базе данных проиндексированных документов:

Данная функциональность может состоять из нескольких этапов:

* + Нахождение документов, соответствующих поисковому запросу;
  + Ранжирование документов в соответствии с их релевантностью поисковым запросам;
  + Кластеризация документов:
    - Целью кластеризации документов является автоматическое выявление групп, семантически похожих документов среди заданного фиксированного множества документов.

**4 слайд:**

Основными критериями качества работы поисковой машины являются:

* Релевантность - степень соответствия запроса и найденного, т. е. уместность результата;
* Полнота индекса;
* Учёт морфологии языка.

**5 слайд:**

Перейдем к классификации машин.

Есть множество способов разбить поисковые машины на различные группы, но сейчас нас интересует разбитие по области поиска:

* Локальные:
  + примером является Elasticsearch;
* Персональные;
* Глобальные:
  + Предназначены для поиска информации по всей сети Интернет либо по значительной её части (Google, Яндекс и др).

Как раз о глобальных поисковых машинах далее пойдет речь.

**6 слайд:**

Поисковые машины Интернета тесто связаны с поисковыми системами. На данном слайде показано распространение различных поисковых систем по странам. Можно заметить, что в основном пользуются поиском Гугл, хотя, в России ситуация сложнее.

**7 слайд:**

Google – самая популярная в мире поисковая система и закономерно обладает самой популярной в мире поисковой машиной – около 90% мирового поиска.

В России ситуация отличается от мировой: здесь сильные позиции удерживает поисковая машина отечественного производства – Yandex захватил более половины рынка поиска в России.

**8 слайд:**

Можно перейти непосредственно к анализу поисковых машин.

Google в работе своей поисковой машине использует следующие алгоритмы:

* PageRank:
  + Числовая величина, характеризующая «важность» веб-страницы;
  + «Вес» страницы А определяется весом ссылки, передаваемой страницей B.
  + По сути, PageRank — это метод вычисления веса страницы путём подсчёта важности ссылок на неё.
* RankBrain:
  + Самообучающаяся система с искусственным интеллектом;
  + Если RankBrain видит слово или фразу, но не знаком с ней, алгоритм может сделать предположение относительно того, какие слова или фразы могут иметь аналогичный смысл и фильтровать результат.
* Hummingbird(обновление):
  + Алгоритм пытается «понять» контекст запроса, а не только отдельные слова.
* Учет персональных данных пользователя.

**9 слайд:**

Yandex в своей поисковой машине использует следующие алгоритмы:

* Индекс качества сайта(ИКС);
  + Включает в себя Тематический индекс цитирования (ТИЦ) - технология поисковой машины «Яндекс», заключающаяся в определении авторитетности интернет-ресурсов с учётом качественной характеристики — ссылок на них с других сайтов. ТИЦ рассчитывается по специально разработанному алгоритму, в котором особое значение придаётся тематической близости ресурса и ссылающихся на него сайтов.
  + В целом это аналог PageRank от Google.
* MatrixNet:
  + Самообучающаяся система с искусственным интеллектом;
  + С помощью MatrixNet можно построить очень длинную и сложную формулу ранжирования, которая учитывает множество различных факторов и их комбинаций;
* Продвинутый анализ морфологии русского языка;
* Учет персональных данных пользователя.

**Слайд 10:**

Учет персональных данных пользователя может иметь как положительные, так и отрицательные стороны. Понятно, что поисковая выдача становится всё более релевантной, но также возникает так называемый «пузырь фильтров».

Негативная сторона персонализированного поиска, явления, при котором веб-сайты определяют, какую информацию пользователь хотел бы увидеть, основываясь на информации о его месторасположении, прошлых нажатиях и перемещениях мыши, предпочтениях и истории поиска.

В результате веб-сайты показывают только информацию, которая согласуется с прошлыми точками зрения данного пользователя. Вся иная информация, как правило, пользователю не выводится.

Так, если у пользователя есть два друга в социальной сети, один делает посты с поддержкой президента, а другой с его критикой, и пользователь делится первыми, а вторые игнорирует, то постепенно в его ленте будут появляться посты со всё более ярко выраженной высокой оценкой президента. Соответственно, он получает намного меньше противоречащей своей точке зрения информации и становится интеллектуально изолированным в своём собственном информационном пузыре.

**Слайд 11:**

Таким образом, сравним поисковые машины Google и Яндекс:

Схожие черты:

* Используют сразу несколько алгоритмов для выдачи релевантных результатов на запрос;
* Используют алгоритмы ранжирования на основе ссылок;
* Используют алгоритмы машинного обучения для выдачи более точных результатов;
* Используют персональные данные пользователей при составлении ответа на поисковой запрос.

Отличительные черты:

* Yandex сильнее , чем Google, учитывает региональную специфику русскоязычных стран.

**Слайд 12:**

Подведем итоги:

* В целом все поисковые машины уходят от алгоритмов ссылочного ранжирования и переходят к алгоритмам машинного обучения для более интеллектуального анализа поисковых запросов;
* Поисковые машины всё чаще используют личные данные пользователя для поднятия точности ответов на поисковые запросы;
* Улучшение алгоритмов поисковых машин может иметь как положительный, так и отрицательный характер.