1. **Опишите процесс составления индекса поисковой машиной**

Составление индекса – индекс — своеобразный словарь, в котором основы упорядочены по алфавиту, а при каждой основе записано, с какой страницы она взята (номер страницы) и на каком месте на этой странице данная основа стояла (номер вхождения).

Структура:

СНОВА /номер страницы + номер вхождения / номер страницы + номер вхождения / номер страницы + номер вхождения /....

1. **Почему индекс поисковой машины называется инвертированным?**

Если в обычном тексте мы идем от страницы к словам, то в индексе поисковая машина идет от слов к страницам, поэтому индекс поисковой машины называется инвертированным или инверсным, то есть обращенным, перевернутым.

1. **Какие поисковые системы хранят проиндексированные страницы полностью?**

* Google
* Yahoo!
* Яндекс

1. **Нужно ли индексировать стоп-слова и цифры?**

Стоп-слова – неважные, служебные слова (предлоги, союзы, сокращения, числа и цифры). Пользователи достаточно часто запрашивают такие слова, поэтому их нужно хранить (особенно для поиска точных цитат, включающих эти служебные слова).

1. **Как поисковая система справляется с высокой флективностью русского языка?**

Решением хранить только основы слов, а окончания и чередования — отбрасывать и хранить отдельно. Это позволяет объединять при поиске разные грамматические формы одного и того же слова, что довольно важно для качества поиска. Для этого поисковик обычно применяет словарь, то есть работает только с известными ему словами. Если все словоформы слова сводятся к его основе на этапе создания индекса, то потом, при задании любой словоформы слова в запросе, она также будет сведена к основе и будут найдены все вхождения этой основы в тексте.

1. **На какие категории можно разделить многообразие запросов пользователя?**
2. **Навигационные запросы** - такие запросы вводят пользователи, которые хотят найти конкретное место в Сети — сайт нужной им компании, сайт интересного им форума, ссылку на блог интересной персоны и т. п.
3. **Информационные запросы** - вводят, когда хотят найти какие-то сведения, информацию в Сети, причем пользователю все равно, где именно она находится.
4. **Транзакционные запросы** – запрос, который вводит пользователь, желающий совершить какое-либо действие в Интернете (купить, заказать, скачать, подписаться или заплатить).
5. **Нечёткие запросы** – общие запросы, не относящиеся ни к одной из вышеперечисленных категорий.
6. **Полнота поиска** - мера того, нашел ли поисковик все нужные веб-страницы, которые есть в Сети.
7. **Точность** - мера качества выданных результатов. Точность вычисляется количество релевантных страниц в общем объеме того, что выдал поисковик.
8. **Какие факторы оказывают значительное влияние на индексацию страниц сайта и дальнейшее ранжирование ссылок на документы в результатах поиска на поисковых системах?**

Можно выделить несколько факторов:

1. Адреса страниц (статистические и динамические, вид адресов страниц).
2. Работа серверных скриптов (заголовки HTTP, дата изменения страницы, сессии и дубликаты, код страниц).
3. **Какой вид должны иметь адреса страниц для лучшего поискового ранжирования?**

Некоторые поисковики, например, Google, при ранжировании страниц учитывают найденные ключевые слова в названии домена, директорий и файлов и придают немного больший вес таким страницам. Значит, для нас важно, чтобы слова в имени страницы были учтены поисковым роботом. Для этого, в качестве знака разделителя между словами, нужно использовать дефис, так как знак подчеркивания не является для поискового робота разделителем слов, так что лучше называть файлы realty-broker-service.html, а не realty\_broker\_service.html.

1. **Какими способами можно влиять на индексацию сайта?**

Для того, чтобы максимально ускорить индексацию сайта, необходимо правильно настроить взаимодействие сайта с поисковыми роботами, и предельно четко им показывать, что нужно сканировать, а что нет.

Способы:

* Файл robots.txt
* МЕТА-тег “Robots”
* If-modified-Since

1. **Составьте план действий на сайте перед добавлением его в поисковые машины используя** [**http://www.ashmanov.com/pap/ivsprep.phtml**](http://www.ashmanov.com/pap/ivsprep.phtml)**.**

План:

* Настроить титульную фразу с помощью тега <title>
* Описать страницу с помощью тега <meta name = “description” content =”>
* Настроить позиции найденных слов в титульной фразе
* Настроить “вес” найденного слова и близость его к началу документа
* Настроить наличие найденных слов в заголовках (h1, h2 и т.д.) и фрагментах, выделенных жирным шрифтом (b, strong)
* Настроить близость найденных слов друг к другу (компактность цитаты)
* По возможности, распространить ссылку своей страницы на других страницах
* Настроить специальный метатег <META name = “ROBOTS”> и файл исключений robots.txt