

За последний год с введением Находки, внутренние факторы стали важнейшими факторами, позволяющими ранжировать сайты документов. Год назад на внутренние факторы никто не обращал внимания, так как по сравнению с внешними факторами влияли не значительно. Однако сейчас значительно возросло их влияние.

Первоначально перед ранжированием документов происходит препроцессинг запроса. Препроцессинг запроса (или «переколдовка») – установка эмпирически подобранных контекстных ограничений. В настоящий момент отмечается тот факт, что контекстные ограничения в Яндексе практически отсутствуют. Полгода назад было серьёзное влияние. Были разнообразные способы переколдовки. И была возможность увидеть как переколдовывается запрос. Сейчас мы сталкиваемся с тем, что таких контекстных ограничений нет. Более того, контент документа и анкор-файлы объединены в одну поисковую зону.

Кроме того, среди нововведений в последнее время можно отметить более широкое добавление к формам слова других слов, таких как однокоренные слова, перевод, транслит, синонимы и расшифровки аббревиатур. В основном встречаются переводы на английский и украинский языки, с другими языками сталкиваться не приходилось. Примером расшифровки аббревиатуры может послужить то, что мы видим, что слово ОСАГО расшифровывается как «обязательное страхование авто гражданской ответственности». Расшифровка аббревиатур была одним из самых первым нововведением расширения ассоциированных слов. Эти расшифровки одно время указывались внизу под формой слова.

Вопрос: влияет ли наличие форм слов ОСАГО, если в тексте страниц есть «страхование»?

Ответ: Здесь влияет именно полная расшифровка. Одно дело, на сколько их сила равна точному вхождению запроса.

Следующий момент при ранжировании – фильтрация по кворуму (определение релевантных пассажей). Многие читали Яндекс на РОМИП 2004, где Илья Сегалович расписал принципы ранжирования поисковой системы. Эта формула работала до 2007 года, пока все могли видеть переколдовку.

Сергей указывает в докладе на то, что при наличии 6 тяжёлых лёгкий кворум уже не проходит. Для прохождения кворума необходимы 6 тяжёлых запросов из 7. Раньше такого не существовало, нужно было меньшее количество тяжёлых слов для прохождения кворума. Это даёт основание полагать, что неполные пассажи практически кворум не проходят. Возможно, Яндекс отошёл от той модели ранжирования, которая освещалась в докладе Яндекса в 2004 году.

Другой важный момент – **текстовое ранжирование**. В теории информационного поиска (informational retrieval) известен такой подход, как $tf \cdot idf$. Что под этим подразумевается? Оценивается вес отдельных слов по отношению к запросу, а затем из суммы этих весов собирается вес страницы. Как правило, idf – логарифм обратный частоте терминов коллекции. Под tf подразумевается нормирование числа вхождений леммы в документ (TF). И зная эту нормировку, мы можем выбирать стратегию расположения числа и плотности ключевых слов при составлении максимально релевантного документа. Выделяется два вида нормировки: 1) по длине документа и 2) по частоте самого частотного термина.

Если рассматривать нормировку по длине документа, то здесь важно количество слов, входящих в документ. То есть получается, что чем длиннее документ, тем ниже релевантность. Перед **оптимизатором** стоит задача сделать как можно больше вхождений терминов в документ, но при этом обеспечить его наименьшую длину, иными словами необходимо наращивать плотность. В таком случае получается, что идеальная плотность стремится к 100%, однако поисковые системы накладывают на такие страницы с подозрительно высокой плотностью специальные санкции. В качестве примера приводится формула из доклада Яндекса на РОМИП-2006:
 $tf = TF_{норм} = TF / (TF + k_1 + k_2 \cdot DL)$, где $k_1 = 1$, $k_2 = 1/350$.

Нормировка по частоте самого частотного термина использовалась в Яндексе в 2006-2007 гг. и была установлена по методике Минюча. Пример: $tf = TF_{норм} = TF / \max(7, TF_{max})$.

Сергей Людкевич предполагает, что в настоящий момент идёт нормировка по длине документа.

Вопрос: тэги входят в длину документа?

Ответ: Смотря какие тэги. Например тег `<title>` учитывается, а `<discription>` - нет.

Вопрос: входят ли стоп-слова в длину документа?

Ответ: да, так как по ним можно искать.

Вопрос: Какие задачи ставить перед [SEO](#)-копирайтером?

Ответ: Реалии таковы, что писать необходимо большие тексты с большим количеством ключевиков.

Поисковые системы по разному учитывают форматирование текста и зоны документа. Тэг <title> в Яндексе имел двойное преимущество перед контентом, то есть употребление слова в этом тэге соответствовало 2-м употреблениям этого слова в контенте. Сейчас эти преимущества должны сохраниться. Однако, если Google и Рамблер полностью учитывает тэг <title>, то Яндекс – только первые 15 слов. Мета-тэги <description> и <keywords> не учитываются ни в одной из вышеназванных ПС. Атрибут alt тега учитывается в Google, в Рамблере – полные слова, входящие в первые 60 символов, а в Яндексе вовсе не учитывается. Что касается URL документа, то в Яндексе и Google он учитывается, а в Рамблере – нет.

В завершении Сергей даёт несколько простых советов:

- 1. Страница должна быть по возможности наилучшим ответом на конкретный запрос.**
- 2. Самые высококонкурентные запросы целесообразно привязывать к главной странице.**
- 3. Идеальный вариант: 1 страница = 1 запрос.**

<http://www.searchengines.ru/>