



МАРКЕТИНГОВАЯ ГРУППА "ТЕКАРТ"

Сергей Людкевич,
руководитель департамента интернет-маркетинга

«Современные аспекты продвижения сайтов в поисковых машинах»

Составление семантического ядра.

Этап 1. Анализ бизнеса клиента и контента его сайта.

Выделение тематических подразделов.
Расстановка приоритетов между ними.

Составление семантического ядра.

Этап 2. Подбор релевантных запросов.

1. Выбор базовых «ядрообразующих» терминов
2. Подбор словочетаний с ними
3. Подбор ассоциативных запросов
4. Анализ запросов с опечатками и ошибками
5. Расстановка приоритетов. Оценка конкурентности.

Составление семантического ядра.

Этап 2. Подбор релевантных запросов.

Инструментарий

<https://ad.rambler.ru/swrds/wrds.pl>

<http://stat.go.mail.ru/stats>

<http://wordstat.yandex.ru/>

Основные факторы, влияющие на релевантность.

1. Текстовое ранжирование
2. Анкорное ранжирование
3. Статический ранг

Текстовое ранжирование.

1. Препроцессинг запроса («переколдовка») – установка эмпирически подобранных контекстных ограничений.

Текстовое ранжирование.

2. Фильтрация по кворуму — определение релевантных пассажей.

Релевантными считаются все полные пассажи и те неполные, сумма весов слов которых превосходитнеобходимый кворум.

Текстовое ранжирование.

Формула кворума Яндекса образца 2004 г.

$$\sum_{i=1}^K P_i \geq QuorumWeight \sum_{j=1}^N P_j$$

$$QuorumWeight = \left(1 - Softness^{\frac{1}{\sqrt{N-1}}} \right)$$

$$P = p^{0,38}$$

$p = ICF = TotalLemms / CF$ — обратная частота термина в коллекции

$TotalLemms$ – общее число вхождений всех лемм в коллекции,

CF – число вхождений леммы в коллекцию.

Текстовое ранжирование.

Подход $tf * idf$.

Учет контекстуального сходства

Учет форматирования текста.

Текстовое ранжирование.

Формула Яндекса образца 2006-2007 гг.

$$tf = TF_{норм} = TF / \max(7, \sqrt{TF_{max}})$$

$$idf = \sqrt{\ln(p)}$$

TF – число вхождений леммы в релевантные пассажи

Текстовое ранжирование.

Привязка запросов из семантического ядра к страницам сайта.

Страница должна быть по возможности лучшим ответом на запрос.

Самые высоконкурентные запросы привязываются к главной странице.

Идеальный вариант: Одна страница — один запрос.

Текстовое ранжирование.

Текстовая оптимизация страниц.

Максимально возможное количество употреблений терминов запроса.

(контент, тег title, выделенные области)

Максимально возможное контекстуальное сходство (словоформы, расположение слов, расстояние между словами)

Анкорное ранжирование.

Формула Яндекса образца 2006-2007 гг.

$$idf = \ln(p)$$

Влияние других терминов анкера

$$P_A = p_A^{0,38}$$

обратная частота по корпусу анкор-файлов.

$$k = P_A / \sum_{i=1}^n P_{Ai}$$

Статические факторы.

1. *PageRank*

Все доноры равноправны.

2. *TrustRank*

Обучающая выборка заведомо «хороших» и заведомо «плохих» доноров.

Противодействие поисковых машин ссылочному спаму.

Непот-фильтр.
«Ты последний»
«Невидимый» индекс.

Выбор наиболее благоприятных доноров.

1. Репутация как поставщика информации.

Реальный сайт.

Полезный авторский контент.

Возраст сайта → max.

Выбор наиболее благоприятных доноров.

2. Репутация как «донора»

Число «сомнительных» исходящих ссылок → min

3. Репутация как «акцептора»

Число «сомнительных» входящих ссылок → min



МАРКЕТИНГОВАЯ ГРУППА "ТЕКАРТ"

Сергей Людкевич,
руководитель департамента интернет-маркетинга

Спасибо за внимание!
Пожалуйста, задавайте вопросы.