




Использование особенностей языка запросов поиска Яндекса для исследований

Трофименко Е.А.

trofimenko.evgeny@rbscorp.ru

Корпорация РБС, начальник отдела исследований и
аналитики





Основные идеи доклада

- Оператор «минус» не применяется к текстам ссылок
- Оператор «~~» вычищает НПС-результаты из выдачи
- Можно узнать, как Яндекс расширяет запрос пользователя
- Контекстных ограничений не существует

Польза:

- Для экспериментов
- Для оценки конкуренции по запросу
- Для поиска альтернативных вариантов продвижения
- Для общей эрудиции



Операторы Яндекса

Присутствие: **+** (плюс) - слово обязано находиться

Исключение:

- ~ ~** (тильда) - исключение в пределах документа (~ предложения)
- (минус) - недокументированный: был исключением в контексте

Возможные контексты поиска:

- Документ (текст), Предложение (текст)
- Ссылки (анкор-файл)
- ...расстояние в несколько слов или предложений, указанное явно операторами **/(-N +N)** или **&&/(-N +N)**



Запрос [слово -слово]: что найдем?

слово - документы, содержащие слово в текстах или во входящих ссылках

-слово - исключаются документы, содержащие слово в текстах

Итог: найденные по ссылкам документы (сниппет м.б. из Я.Каталога)

The screenshot shows a web browser window with the Yandex search engine. The address bar contains the URL: <http://yandex.ru/yandsearch?text=слово+-слово&stpar2=%2Fh1%2Ftm8%2F>. The search bar contains the text "слово -слово" and a "Найти" button. Below the search bar, there are options for "в найденном" and "в Москве", and a link for "расширенный поиск". The search results show 150 thousand pages found. The first result is from mignews.com.ua, with a snippet: "найден по ссылке: Новое русское слово" (Европа-СНГ)...". The second result is <http://wordstat.yandex.ru/>.

Для чего нужны найденные по ссылке («НПС») результаты с фиксированными словами?



Для проведения экспериментов:

- Выбор множества ссылок на разные документы, которые содержат нужные наборы слов или их последовательности
- Анализ «весов» конкретных слов в текстах ссылок («анкор-файле»)
- Прикидка весов ссылок



~ и ~~ : форсируем поиск по текстам и удаляем НПС-результаты

Добавляя в любой запрос исключение ~~абракадабры, удаляем НПС:

Для длинных запросов доля найденных по ссылке результатов
относительно высока:

авто	~0.35% НПС
купить авто	~5% НПС
купить авто москва	~13% НПС
купить подержанное авто москва	~25% НПС



Расширение пользовательских запросов

Яндекс и раньше мог добавлять в запрос новые слова («что такое X»), но делал это редко, индивидуально.

Сейчас - расширение запроса поставлено на поток.

- Переходы из одной части речи в другую (*гостиницы в Москве -> московские гостиницы*)
- Транслитерация («*mazda*» -> «мазда»)
- Аббревиатуры (*МГУ -> Московский государственный университет*)

Как узнать слова, которыми расширяется запрос?

Используем операторы исключения.

Исключаем точную форму слова: оставляем переформулировки



При исключении слова из запроса - в выдаче остаются и подсвечиваются переформулировки (+найденное в URL):

слова запроса ~~слово	Перестало работать: исключает все формы, и переформулировки тоже
слова запроса –слово	Работает, но оставляет смесь переформулировок и НПС
запрос ~~!(слово слова) запрос ~~!!(слово)	Работает.

Изменение «колдунщика»: отмена ограничения расстояний



Колдунщик = расстановка неявных для пользователя ограничений на расстояние между словами, известен с 2004.

Например, для запроса «новый год» находились документы, содержащие от «год новый» до «новый [*] [*] год»

Эти ограничения можно было посмотреть. Но это закончилось. Однако вручную введенные ограничения обрабатывали.

Как узнать реальные ограничения на расстояние между словами?

Попробуем подобрать...

Пытаемся подобрать: перебор 7 частотных операторов



По статистике запросов Корпорации РБС, наиболее часто использовались:

&	Относительно: 100%	в пределах одного предложения
&/(-2 4)	9%	-2 +4 слов
&/(-1 3)	10%	-1 +3 слов
&/(1 1)	2%	строго по порядку
&&/(-7 7)	15%	в пределах 7 предложений
&&/(-3 3)	15%	3 предложений
&&	7%	в пределах документа

Перебор вариантов НЕ ДАЕТ РЕЗУЛЬТАТОВ...

А существуют ли теперь в Яндексе контекстные ограничения?



Из релиза Яндекса, Магадан:

«Мы смягчили фильтрацию отбора документов для ранжирования, что привело к улучшению ранжирования по запросам, для которых релевантные документы содержат слова запроса далеко друг от друга»

Попробуем поискать эти ограничения.

1. Берем запрос (+скачать +реферат), оба слова должны находиться
2. Исключаем результаты поиска, в которых слова находятся «слишком близко» (от 1 слова до 10 тыс. предложений):
~~(+скачать & +реферат)
3. Смотрим, как меняется число найденных документов... Надеемся, что оно станет нулевым тогда, когда расстояние совпадет с расстоянием в заколдованном запросе...

Кажется, ограничений не существует... находятся документы на любых расстояниях...



(+скачать +реферат) ~~(+скачать [ОПЕРАТОР] +реферат)





Итого:

- Можно искать «найденные по нужной ссылке» результаты
- Можно узнать, как Яндекс переформулирует запрос пользователя
- Можно отказаться от представлений о контекстных ограничениях в Яндексе...

Трофименко Е.А.

trofimenko.evgeny@rbscorp.ru

Корпорация РБС, начальник отдела исследований и аналитики

ВСЁ!!! 😊