



**Web Effector**



**Региональное ранжирование в эпоху MatrixNet**

Сергей ЛЮДКЕВИЧ, начальник отдела аналитики

# ОБУЧАЮЩАЯ ВЫБОРКА



Набор запросов  $q^{(i)}$

Набор документов  $d_j^{(i)}$  для каждого запроса  $q^{(i)}$

**$Rel(q^{(i)}, d_j^{(i)})$**  - ручная оценка соответствия документа запросу  
(выполняется ассессорами)

Конкурс «Интернет-математика – 2009»:

**$Rel(q, d)$**  - значения из диапазона **[0, 4]**

(4 – «высокая релевантность», ..., 0 – «нерелевантно»)

# ФАКТОРЫ РАНЖИРОВАНИЯ



Набор факторов ранжирования

$$F = (f_1(q,d) , \dots , f_N(q,d))$$

Базовые факторы – **420** (апрель 2010)

Производные факторы (произведения базовых факторов) –  
**несколько тысяч**

# ФУНКЦИЯ РЕЛЕВАНТНОСТИ



Числовое соответствие документа запросу

$$\text{Fr}(\mathbf{q}, \mathbf{d}) = \text{Fr}(\mathbf{F}(\mathbf{q}, \mathbf{d})) = \text{Fr}(f_1(\mathbf{q}, \mathbf{d}), \dots, f_N(\mathbf{q}, \mathbf{d}))$$

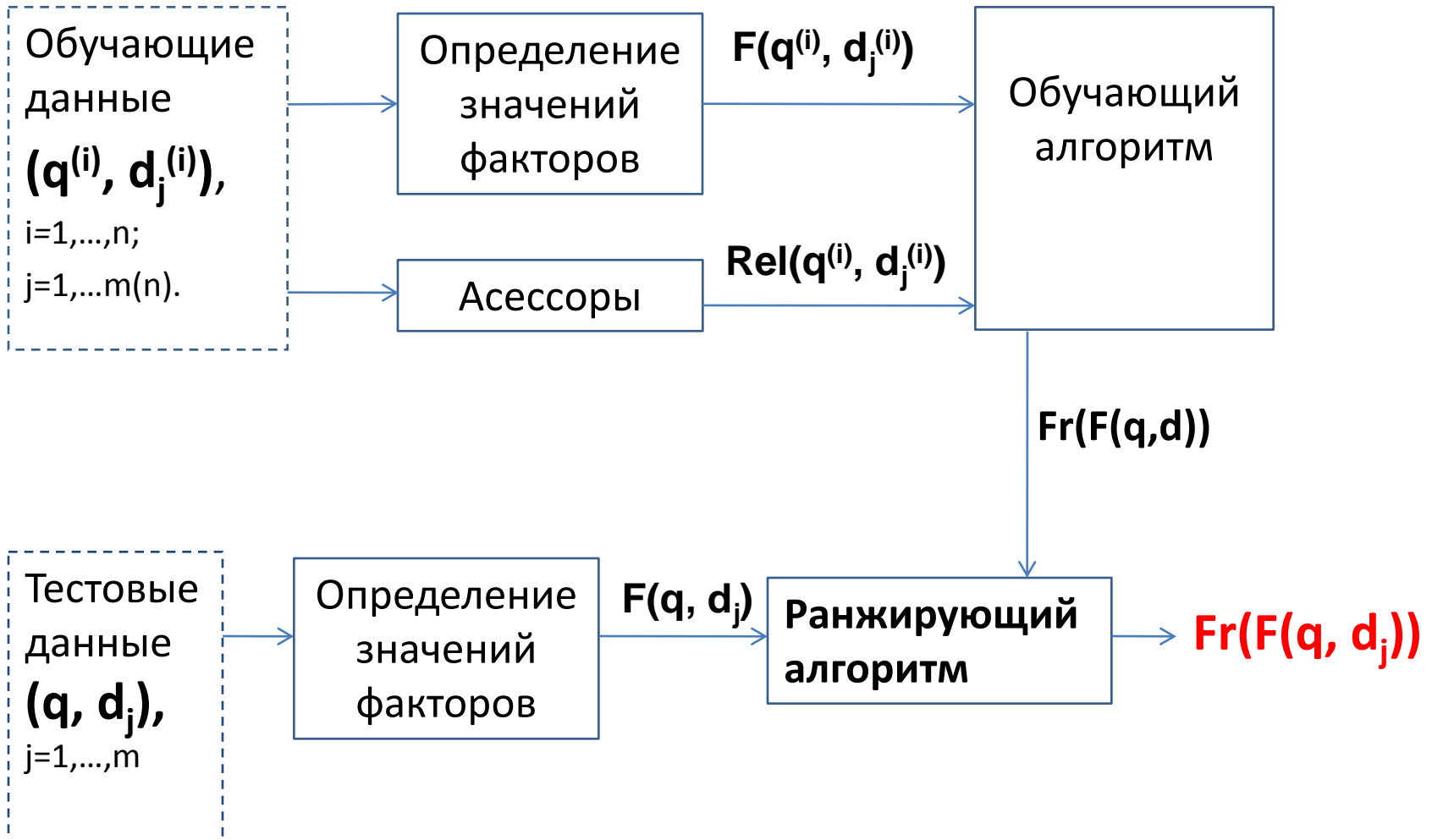
Метод построения функции релевантности в релизе «Снежинск»:

Жадный (greedy) алгоритм **MatrixNet**

Подробнее о MatrixNet:

<http://company.yandex.ru/technology/matrixnet/>

# МАШИННОЕ ОБУЧЕНИЕ



# РЕГИОНАЛЬНЫЕ ФОРМУЛЫ



Отдельные функции релевантности:

- Москва
- Санкт-Петербург + Екатеринбург + 16 городов России релиза «Арзамас+16»: Новосибирск, Самара, Воронеж и др.
- Украина
- Белоруссия
- Казахстан
- Общероссийская + все остальные города России + остальные зарубежные страны + выдача без учета региона (Земля)

# ПРИМЕРЫ ФАКТОРОВ РАНЖИРОВАНИЯ



## Запросные

- длина запроса в словах;
- коммерциализированность запроса.

## Текстовые

- наличие точного вхождения запроса в тексте документа;
- наличие точного вхождения запроса в заголовке документа;
- $tf*idf$ ;
- различные модификации формулы Okapi<sub>BM25</sub>.

# ПРИМЕРЫ ФАКТОРОВ РАНЖИРОВАНИЯ



## Ссылочные статические

- PageRank;
- ТИЦ;
- количество ссылок на документ.

## Ссылочные динамические

- процент ссылок на документ, содержащих точное вхождение запроса;
- $tf*idf$  по текстам ссылок.

# ПРИМЕРЫ ФАКТОРОВ РАНЖИРОВАНИЯ → → → →



## **Свойства документа и сайта**

- это важный сайт (Wikipedia);
- это главная страница сайта;
- порнографичность контента.

## **Антиспамовые**

- содержит ли попандеры или кликандеры;
- естественность текста;
- уникальность контента.

# ПРИМЕРЫ ФАКТОРОВ РАНЖИРОВАНИЯ → → → →



## Поведенческие

- посещаемость сайта;
- кликабельность документа в выдаче.

## Региональные

- совпадение региона запроса и документа;
- региональность ссылок на документ;
- употребление названия региона в тексте документа и в тексте ссылок на документ.

# УЛУЧШЕНИЕ РЕГИОНАЛЬНЫХ ФАКТОРОВ РАНЖИРОВАНИЯ



- Присвоение сайту данного региона  
<http://help.yandex.ru/webmaster/?id=1074582>
- Получение входящих ссылок с сайтов данного региона
- Употребление названия данного региона в тексте документа
- Употребление названия данного региона в текстах входящих ссылок на документ

Спасибо за внимание!



Для продолжения темы посетите

[www.webeffector.ru](http://www.webeffector.ru)

115191, Россия, Москва,

ул. Б. Тульская, д. 13, 4-й этаж ТЦ «Ереван Плаза»

Телефон: (495) 772-97-91 (многоканальный)