

# ТЕМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ

# Тематическая классификация

Зачем?

- релевантность (в т.ч. рекламы)
- вертикальные поиски

Данные

- список запросов
- коллекция текстов
- веб
- клики

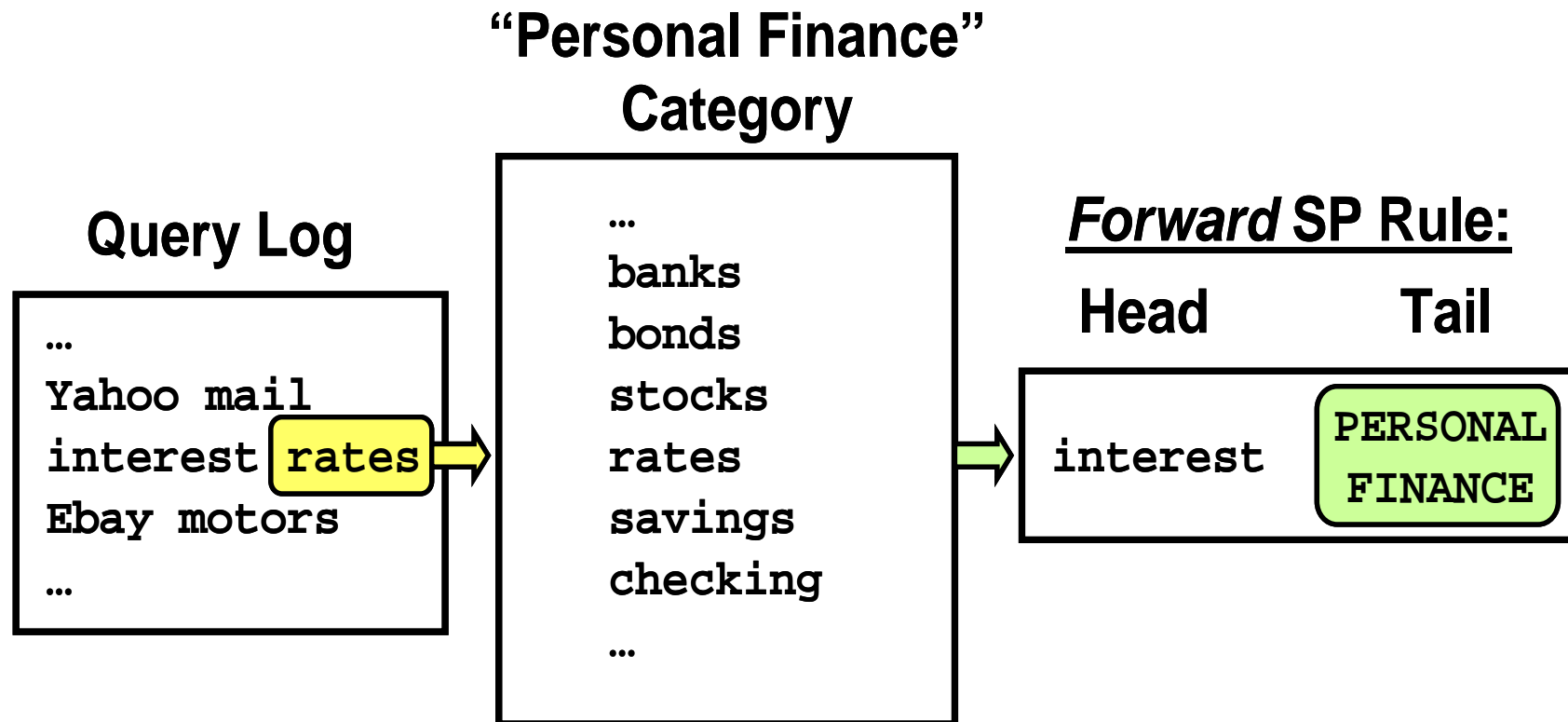
# Selectional Preferences: Step 1

- Obtain a large log of unlabeled web queries
- View each query as pairs of lexical units:
  - <head, tail>
  - Only applicable to queries of 2+ terms
  - Queries with  $n$  terms form  $n-1$  pairs
  - Example: “directions to DIMACS” forms two pairs:
    - <directions, to DIMACS> and <directions to, DIMACS>

# Selectional Preferences: Step 2

- Obtain a set of manually labeled queries
- Check the heads and tails of each pair to see if they appear in the manually labeled set
- Convert each  $\langle \text{head}, \text{tail} \rangle$  pair into:
  - $\langle \text{head}, \text{CATEGORY} \rangle$  (*forward preference*)
  - $\langle \text{CATEGORY}, \text{tail} \rangle$  (*backward preference*)

# Selectional Preferences: Step 2



# Selectional Preferences: Step 3

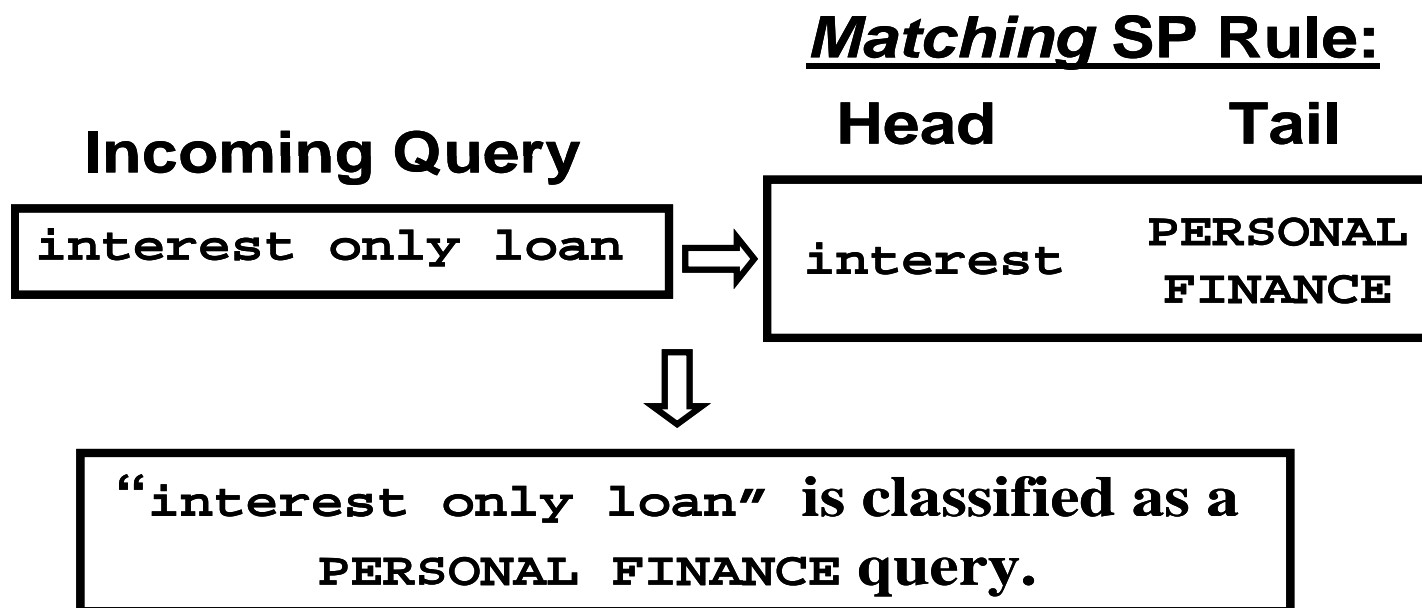
- Score each preference using Resnik's Selectional Preference Strength formula:

$$\begin{aligned} S(x) &= D(P(U|x) || P(U)) \\ &= \sum_u P(u|x) \log_2 \left( \frac{P(u|x)}{P(u)} \right) \end{aligned}$$

- Where  $u$  represents a category, as found in Step 2.
- $S(x)$  is the sum of the weighted scores for every category associated with a given lexical unit

# Selectional Preferences: Step 4

- Use the mined preferences and weighted scores from Steps 3 and 4 to assign classifications to unseen queries



# Selectional Preference Rule Examples

## Forward Rules

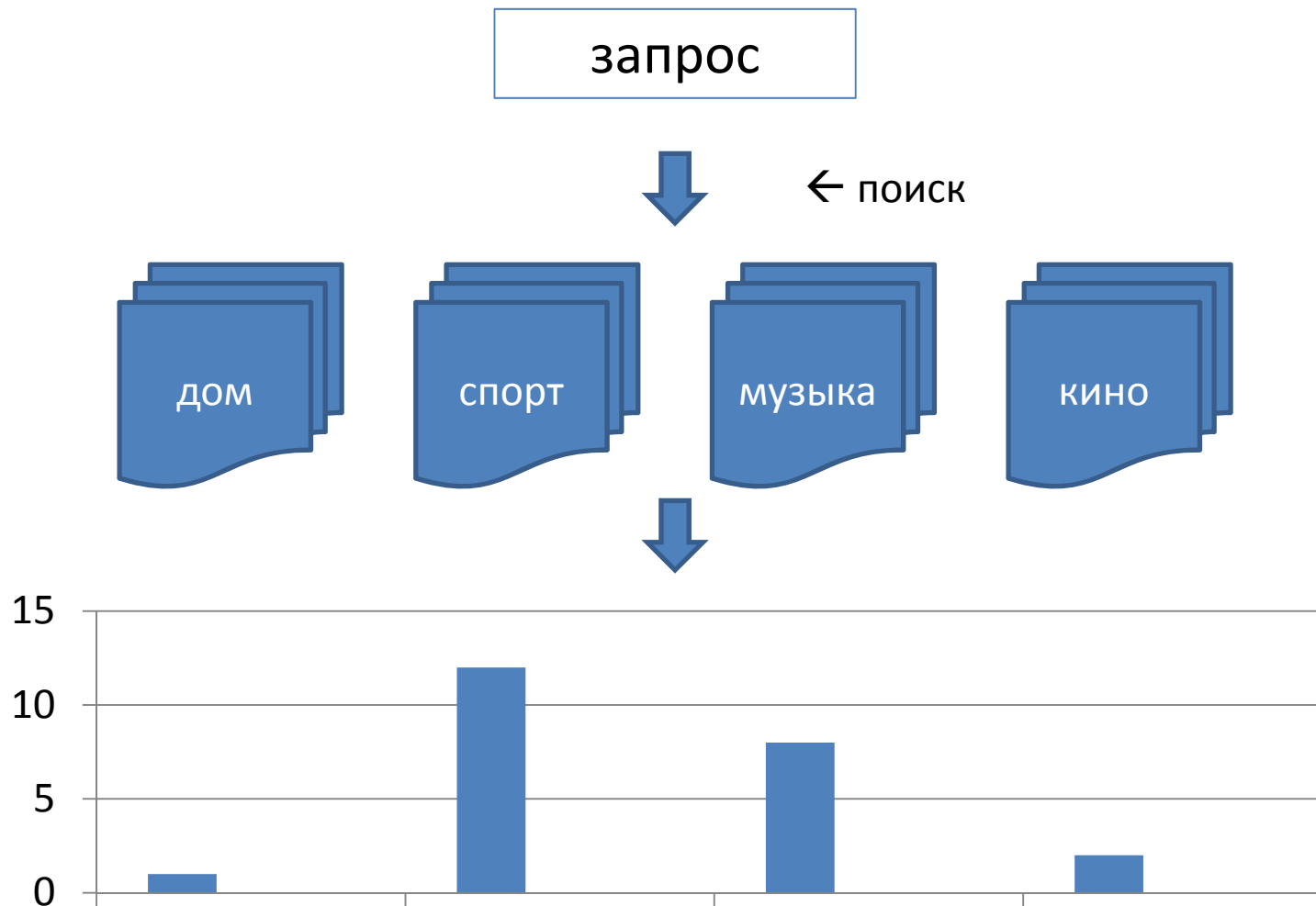
- harlem club **X**
  - ENT->0.722
  - PLACES->0.378
  - TRAVEL->1.531
- harley all stainless **X**
  - AUTOS->3.448
  - SHOPPING->0.021
- harley chicks with **X**
  - PORN->5.681

## Backward Rules

- **X** gets hot wont start
  - AUTOS->2.049
  - PLACES->0.594
- **X** getaway bargain
  - PLACES->0.877
  - SHOPPING->0.047
  - TRAVEL->0.862
- **X** getaway bargain hotel and airfare
  - PLACES->0.594
  - TRAVEL->2.057



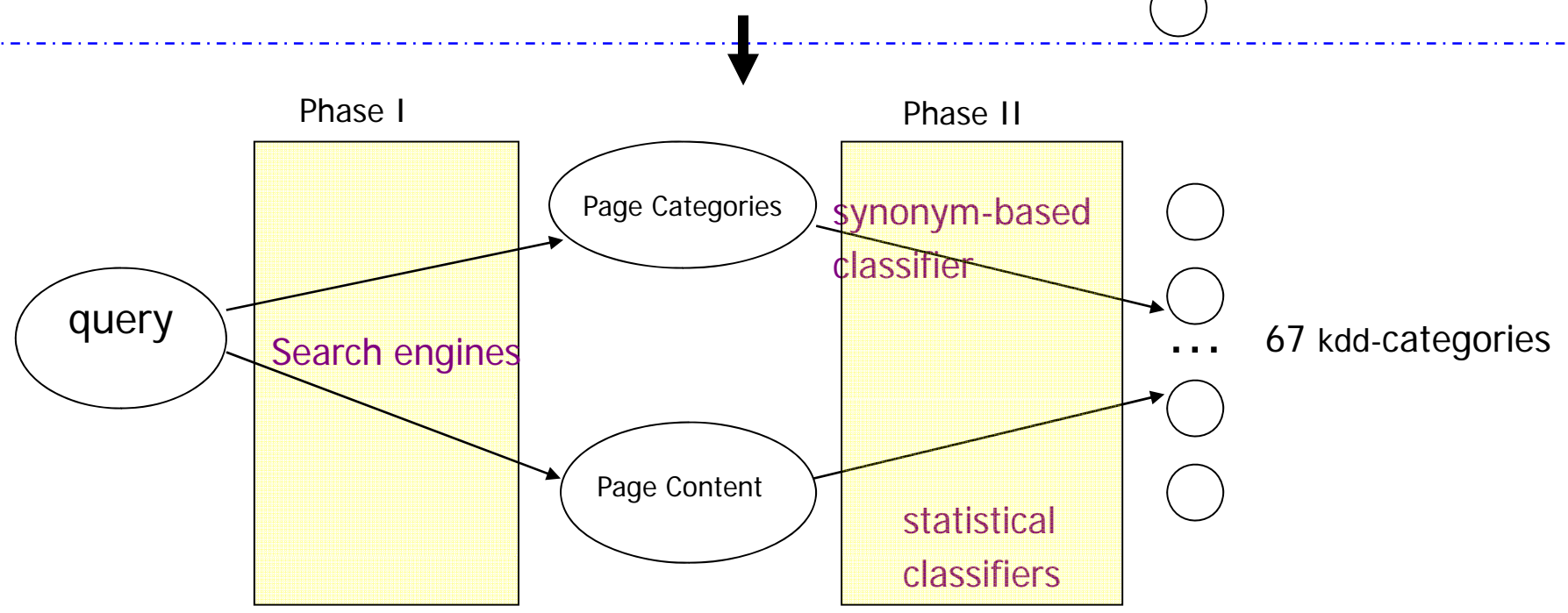
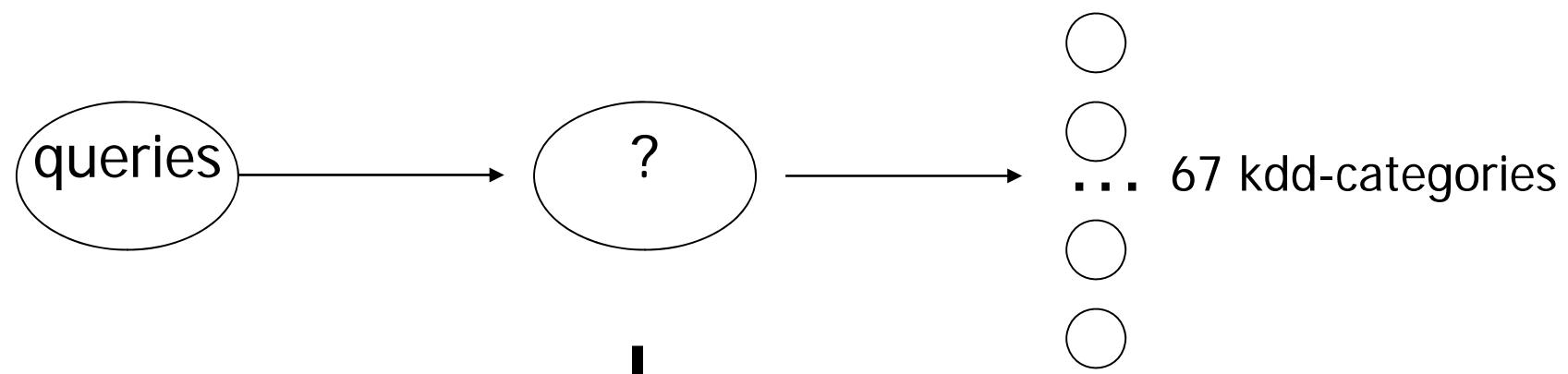
# На основе коллекции текстов



# KDD Cup 2005

- Классификация запросов
- 800,000 запросов, 67 категорий
- есть примеры, нет обучающего множества
- нет подробного описания категорий
- ответ системы: до 5 категорий
- оценка: 800 запросов оцениваются тремя ассессорами
- метрика: F1

# Nature of Problem



# Phase I: From queries to pages and categories

- Input:
    - A query:  $Q_i$ ,
  - Output:
    - $\langle \text{Page list}_i, \text{Category list}_i \rangle$
  - Approach:
    - through Search Engines (SE)
- We collected
    - 40 million entries
    - 50GB
  - Search engines
    - Lumur (CMU open source)
    - Google
    - ODP
    - Looksmart

## [Computerworld](#)

Category: [Computers](#) > [News and Media](#) > [Magazines and E-zines](#) > [Magazines](#)

News and product coverage for information technology managers.

[www.computerworld.com/](#) - [Cached](#) - [Similar pages](#)

## [Computer Associates](#)

Category: [Computers](#) > [Security](#) > [Authentication](#) > [Single Sign-On](#)

Makers of eTrust Single Sign-On. Automates access to authorized Web services and enterprise applications.

[www.ca.com/](#) - [Cached](#) - [Similar pages](#)

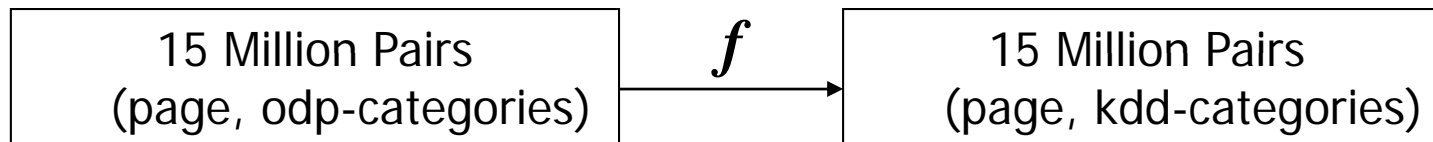
# Phase II.a Synonym-based Classifier: using directories

- 67 KDD-categories in KDDCUP
- ↕
- 172,565 in ODP/Google,  
272,405 in Looksmart

- For each of the KDDCUP category
  - Apply Wordnet to find the corresponding synonyms in the categories of ODP (Google) and Looksmart, respectively
- This produces one mapping function  $f$  for each directory
  - Also returns a rank by matching frequency
- Advantage
  - Fast,
  - Precise
- Disadvantage
  - Many of the 172K and 272K categories from ODP/Google and Looksmart do not map to KDDCUP categories
  - This may result in low recall

# Phase II.b: Statistical Classifiers

- Statistical Classifiers
  - Support Vector Machine (SVM): mapping pages to KDDCUP categories
- Training Data
  - 15 million pages with categories from ODP Directory
  - Apply the mapping  $f$  from Phase II.a, to build training data.



- Application of the classifier
  - Construct a virtual document for each query by combining the snippets from the returned pages given in Phase I.
  - Classifier returns category and rank

# Component Classifier Integration

- We follow an *ensemble learning* approach
  - Each classifier returns the category and rank
  - The two kinds of classifiers have the similar performance.
  - We integrate the different classifiers together by a weighted sum of the ranks
    - Weights can be determined by *validation data set*:
      - Based on the performance on the 111 sample data;
      - Assign different weight values for a classifier on different categories
      - The higher the precision, the higher the weight value
    - We have also tried to use equally weighted component classifiers

# Final Result Generation

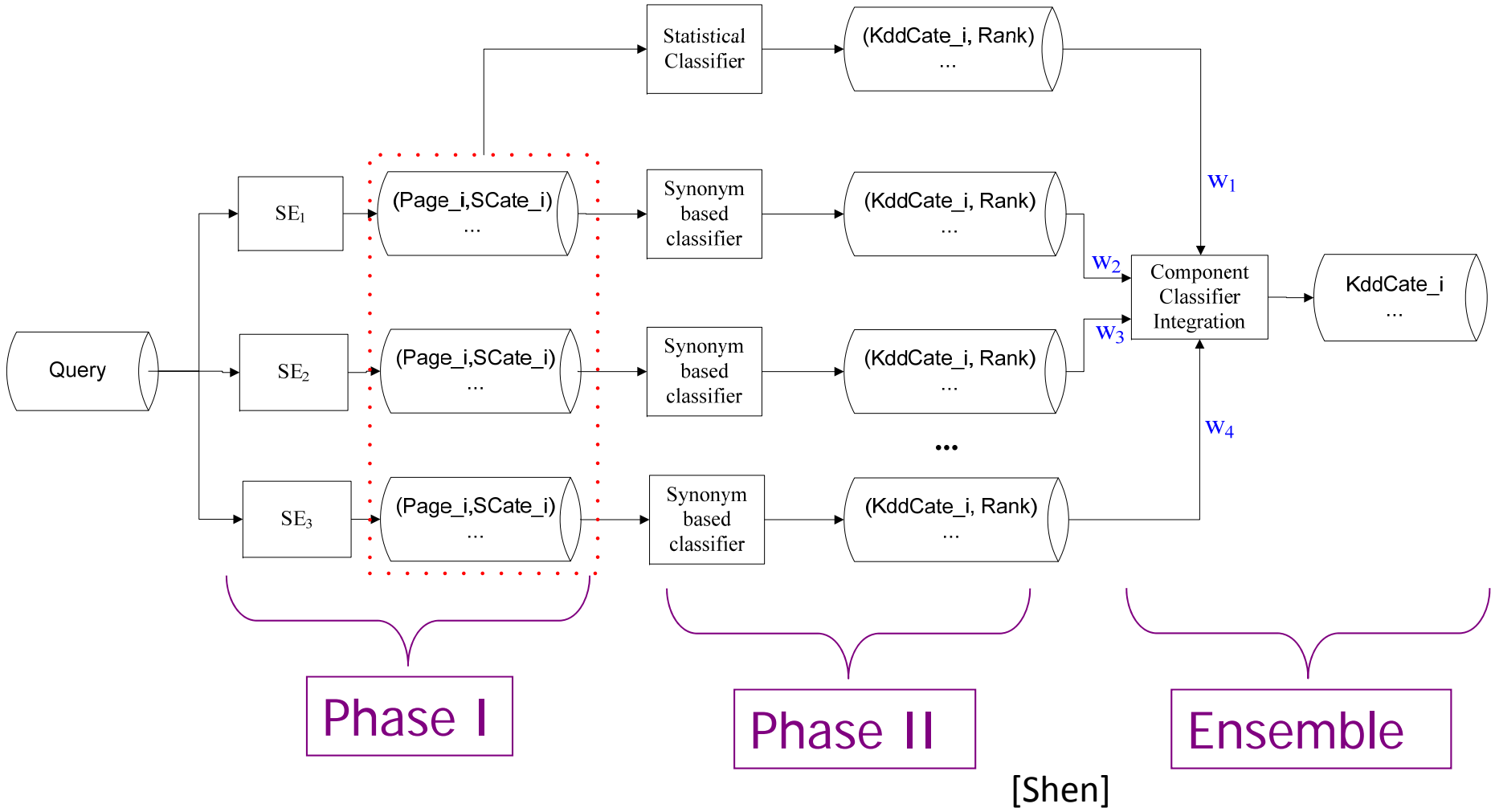
- Two Solutions: One for each evaluation criteria
  - **S1**: Using the validation data set is expected to achieve better precision measure
    - Since each component classifier is highly weighted on the classes where it achieves high precision.
  - **S2**: Equally weighted combination is expected to achieve higher F1 performance
    - Since the recall is relatively high
  - Evaluation Results (<http://www.acm.org/sigs/sigkdd/kdd2005/kddcup.html>)

	Submission ID	Precision	F1
<b>S1</b>	37	0.423741	0.426123
<b>S2</b>	22	0.414067	0.444395

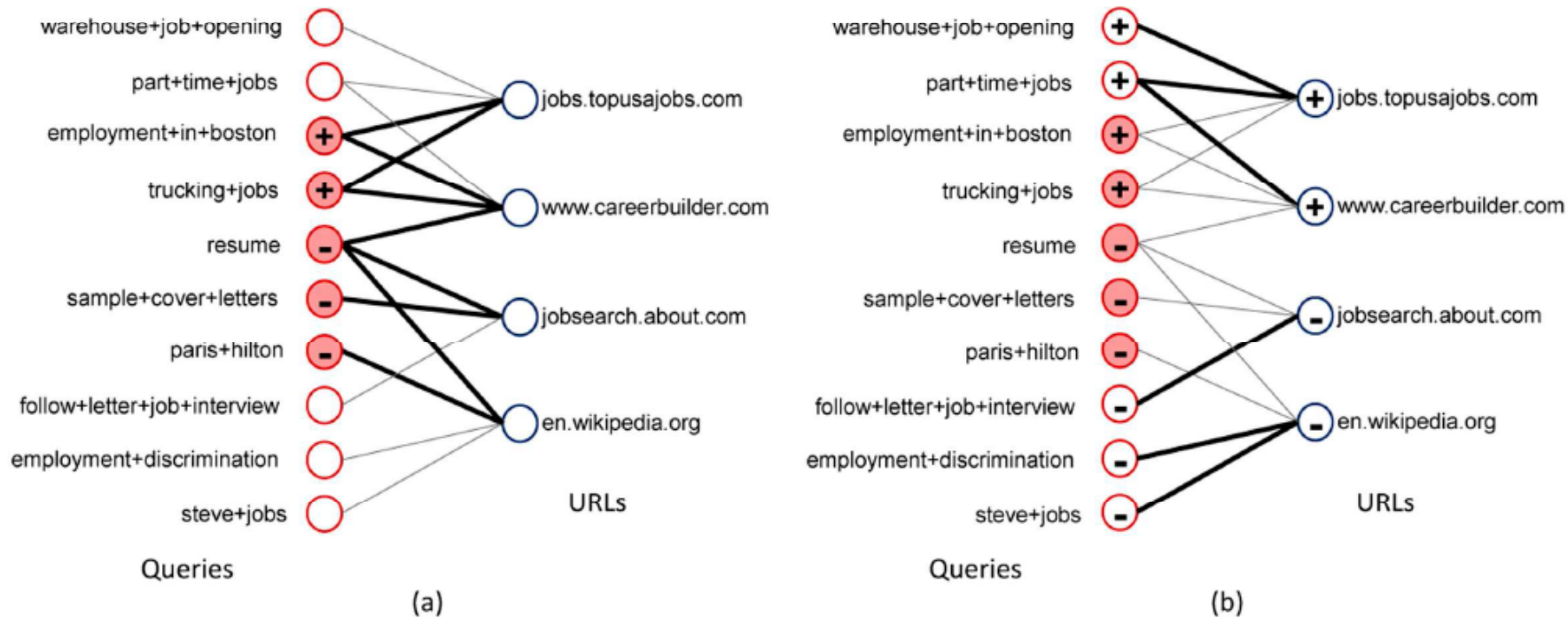
- The Results are generated automatically.



# Putting them together



# Классификация на основе кликов



1. Расширение обучающего множества
2. Классификация на основе лексических признаков
3. Комбинация методов

Li et al. SIGIR2008

# Классификация запросов

$$p_{\lambda}(y|x) = \frac{\exp^{\sum_j \lambda_j \phi_j(x,y)}}{\sum_y \exp^{\sum_j \lambda_j \phi_j(x,y)}}$$

$x$  – запрос,  $y$  – класс (0/1)

Признаки  $\phi(x, y)$  -  $n$ -граммы

[britney spears] →

britney, spears,

<s> britney, britney spears, spears </s>,

<s> britney spears, britney spears </s>

# Распространение меток

- $W$  – матрица  $m \times n$ ,  $w_{ij}$  – количество кликов на документ  $j$  по запросу  $i$
- $F$  – матрица  $m \times 2$ ,  $f_{iy}$  – вероятность принадлежности запроса  $i$  классу  $y$
- $F^0$  – первоначальная разметка  
 $B = D^{-1/2}W$

- Итерации:

$$H^i = B^T F^{i-1};$$
$$F^i = \alpha B H^i + (1 - \alpha) F^0$$

# Комбинация

**Input:** matrix  $F^0$  and matrix  $B = D^{-1/2}W$

**Output:**  $F^*$  and  $\lambda^*$

- 1: Initialize  $F^* = F^0$ , and initialize  $\lambda$  as random;
- 2: **repeat**
- 3: Find  $\lambda^* = \underset{\lambda}{\operatorname{argmin}} Q(F^*, \lambda)$  using stochastic gradient descent;
- 4: Find  $F^* = \underset{\lambda}{\operatorname{argmin}} Q(F, \lambda^*)$  using Algorithm 1, where the input are  $F^c(\lambda^*)$  and  $B$ ;
- 5: **until** the value  $Q(F^*, \lambda^*)$  converges

# БЛИЗКИЕ ЗАПРОСЫ

*михаил булгаков  
мастер и маргарита  
мастер и маргарита фильм  
владимир бортко  
тарас бульба фильм*

*богдан ступка  
михаил боярский  
д'артаньян  
три мушкетера  
александр дюма*

# Подробнее

1. Уточнение:  
*золотое кольцо → золотое кольцо с бриллиантом*
2. Снятие неоднозначности:  
*ягуар → ягуар животное*
3. Расширение:  
*золотое кольцо → ювелирные украшения*
4. Синоним:  
*японская вишня → сакура*



# Еще подробнее

5. Другой запрос на ту же тему:  
*мерседес → ауди*
6. Смена поисковой цели:  
*купить санки → детский мир*
7. Перевод:  
*коралловый клуб → coral club*

# Близкие, но «про другое»

- Опечатки: *курсовая робота – курсовая работа*
- Транслитерация: *золото – zoloto*
- Раскладка: *lbvf ,bkfy – дима билан*
- Реникса: *otbertka – отвертка*

# Хороший запрос – это не просто

*Назовите глагол из вопроса, помещенного  
на борту транспортного средства  
подопечных Фатиха Терима  
на первенстве континента?*

IX Кубок Яндекса по поиску (2008)

<http://kubok.yandex.ru>

# Работа мысли

- Фатих Терим
- сборная турции на борту самолета
- сборная турции "на борту" самолета
- сборная турции надпись "на борту" самолета
- сборная турции надпись "на борту" автобуса
- сборная турции надпись на автобусе
- футбол "сборная турции" надпись на автобусе
- чемпионат европы футбол "сборная турции" надпись на автобусе

# Ответ

**Вместит ли автобус всю страсть Турции?**



harry potter

Search

Options

harry potter movie  
 harry potter 7  
 ...potter and the half blood...  
 harry potter 6  
 ...potter and the deathly hal...

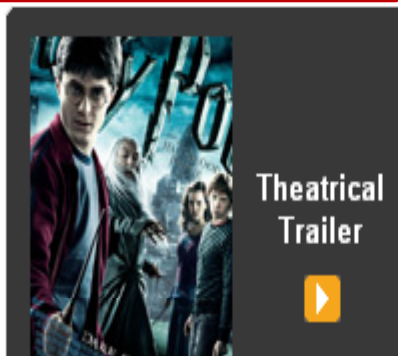
Explore related concepts:

J. K. Rowling	daniel radcliffe
Wizarding World	Professor
Deathly Hallows	<b>Harry Potter books</b>
MuggleNet	Goblet of Fire

Search Pad

SearchScan - On

304,000,000 results for harry potter:



### Harry Potter and the Half-Blood Prince (2009)

[movies.yahoo.com](http://movies.yahoo.com)

[Yahoos B+](#), [Critics B+](#) Voldemort is tightening his grip on both the Muggle and wizarding worlds and... [more](#)

Running Time: 2 hrs 33 mins


Ask.com Search results for "chevrolet".

Search bar:  Search

Navigation: Web Images News Deal\$ Videos Q&A Beta More

Showing 1-10 of 7,870,000 results

**Deals** [Today's best deals for chevrolet](#)

**Chevrolet**  
  
**CHEVROLET**  
[Source](#)

**Chevrolet** sells an impressively wide range of vehicles, from subcompact hatchbacks to huge vans and SUVs. If you're looking for a... [More »](#)  
See Also: [Official Site](#) · [Chevrolet Models](#) · [Reviews](#) · [History](#)  
Search For: [Chevrolet Accessories](#) · [Images](#)  
Popular Models: [Silverado 1500](#) · [Camaro](#) · [Corvette](#) · [Malibu Hybrid](#)

[Продаешь Шевроле?](#) Sponsored Result  
Размести бесплатное объявление На новой торговой площадке Рунета  
[Avito.ru](#)

[2010 Chevy Corvette | Performance Cars | Corvette Cars](#)  
[Chevrolet Home ... Experience Chevrolet ... Close Chevrolet Worldwide...](#)  
[www.chevrolet.com/corvette/](#) Cached

**Related Searches**

- [Chevrolet Camaro](#)
- [Chevrolet History](#)
- [Chevrolet Parts](#)
- [Chevrolet History Beginning](#)
- [Chevrolet Logo](#)
- [Chevrolet History of Cars](#)
- [Who Invented Chevrolet](#)
- [Chevrolet Truck Parts](#)
- [GMC](#)
- [Pontiac](#)
- [General Motors](#)
- [Louis Chevrolet](#)

Интернет | Изображения | Видео | Дополнительно | MSN | Hotmail

bing  
Beta

chevrolet

Показать все  Только на указанном языке: русский язык  Только из ...

ВСЕ РЕЗУЛЬТАТЫ

**СВЯЗАННЫЕ  
ПОИСКОВЫЕ ЗАПРОСЫ**

- Chevrolet Aveo
- Chevrolet Lacetti
- Chevrolet Niva
- Chevrolet Epica
- Chevrolet Viva
- Chevrolet Украина
- Epica**
- Chevrolet авто салон  
сочи

**Автомобили Chevrolet!** · [www.avito.ru](http://www.avito.ru)  
Chevrolet новые и б/у. Покупка и продажа по выгодной цене!

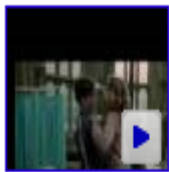
**Chevrolet - Шевроле - купить авто ...**  
Chevrolet Шевроле купить авто Шевролет [chevrolet.ru](http://chevrolet.ru) Chevrolet дилер официальный дилер Chevrolet дилеры ...  
[www.chevrolet.ru](http://www.chevrolet.ru) · [Кэшированная страница](#)

**Шевроле Автомир - официальный ...**  
Компания АВТОМИР - официальный дилер Chevrolet. Она одной из первых открыла официальный сайт ...  
[chevrolet-avtomir.ru](http://chevrolet-avtomir.ru) · [Кэшированная страница](#)

Автомобили Chevrolet Lacetti, Aveo



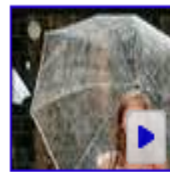
Video results for **hermione granger**



[Hermione Granger](#)

3 min 23 sec - 4 May 2006

[www.youtube.com](http://www.youtube.com)



[Emma Watson/Hermione Granger \\*Lucky\\*](#)

3 min 26 sec - 28 Jan 2008

[www.youtube.com](http://www.youtube.com)

Searches related to **hermione granger**

[hermione granger pictures](#)

[ginny weasley](#)

[harry potter quiz](#)

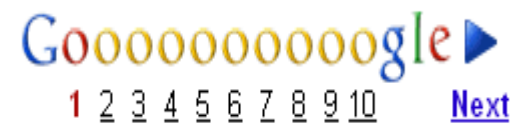
[hermione granger costume](#)

[ron weasley](#)

[draco malfoy](#)

[harry potter](#)

[emma watson](#)



hermione granger

[Search within results](#) - [Language Tools](#) - [Search Help](#) - [Dissatisfied? Help us im](#)

гарри поттер

в найденном

Любые [Обои](#) [Большие](#) [Средние](#) [Маленькие](#) [Портреты](#)



Поищите также:

[гермиона](#) [harry potter](#) [джинни уизли](#)  
[рон уизли](#) [дамблдор](#) [драко малfoy](#)  
[хогвартс](#) [снейп](#) [ден рэдклифф](#)

▶ [Яндекс.Видео](#)



Гарри Поттер 3

[Еще по запросу](#)  
[«гарри поттер»](#) 71145



Гарри Поттер и Орден феникса  
(Harry Potter and the Order of the...)

240×215 ... 699×631  
[www.kino-govno.com](http://www.kino-govno.com)



73. Гарри Поттер 5 / Гарри Поттер 5

Просмотров  
62×100 ... 850×1360  
[gallery.galaxy-portal.ru](http://gallery.galaxy-portal.ru)



Гарри Поттер и орден Феникса  
/ Harry Potter and the Order of the...

120×90 ... 1500×1125  
[relax.grad66.ru](http://relax.grad66.ru)

эмма уотсон

в найденном

Любые [Обои](#) [Большие](#) [Средние](#) [Маленькие](#) [Портреты](#)



Поищите также:

[руперт гринт](#)

[ден рэдклифф](#)

[том фелтон](#)

[бонни райт](#)

[эмма робертс](#)

[званна линч](#)

[emma watson](#)

[эмма уотсон 2009](#)

[Яндекс.Фотки](#)



[Еще по запросу:](#)  
[«ЭММА УОТСОН»](#) 1441



**Эмма Уотсон (Emma Watson)** in  
Teen Vogue (фотосессия.  
383×511 ... 500×666  
[www.port2all.dp.ua](http://www.port2all.dp.ua)



**Эмма Уотсон (Emma Watson)** in  
Teen Vogue (фотосессия.  
511×383 ... 690×518  
[www.port2all.dp.ua](http://www.port2all.dp.ua)



**Эмма Уотсон (Emma Watson)** in  
Teen Vogue (фотосессия.  
511×391 ... 700×535  
[www.port2all.dp.ua](http://www.port2all.dp.ua)

# Источники данных

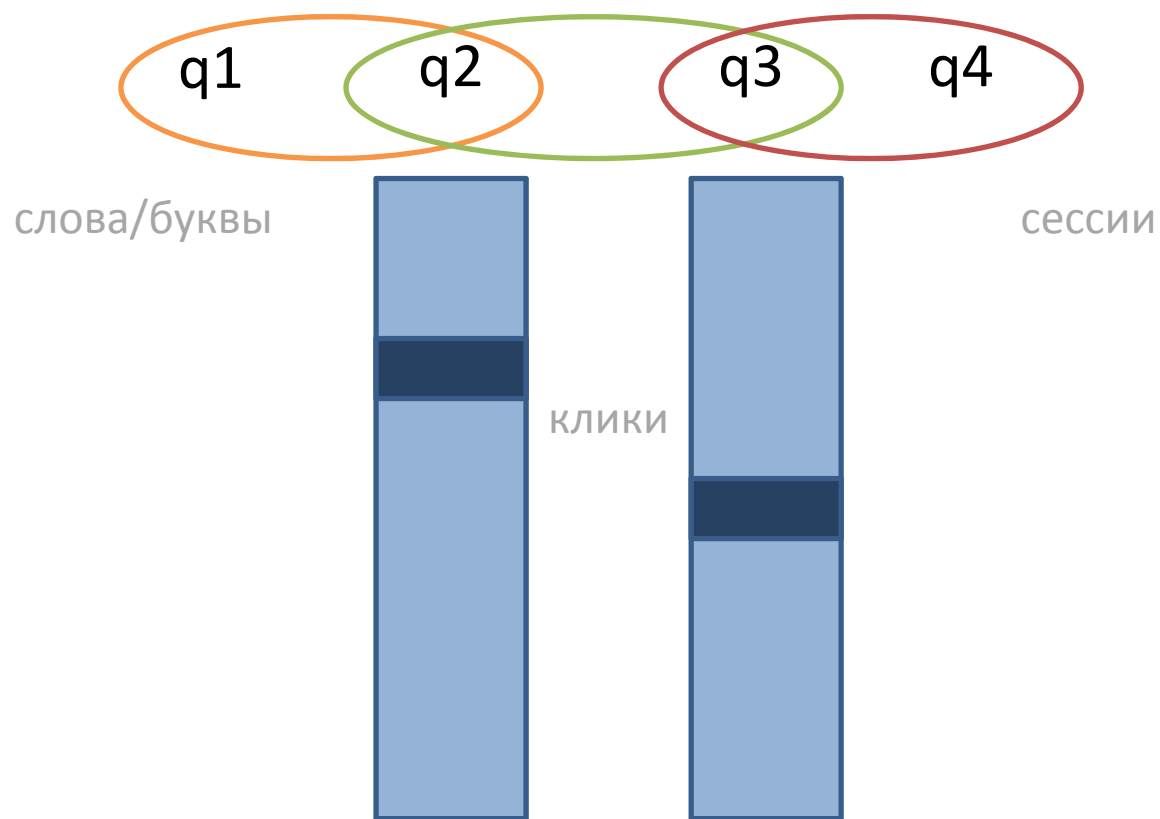
- Лог запросов
- Текст ссылок
- Корпус текстов

*ford* → *ford focus, ford fusion, ford mondeo*

*карта* → *карта памяти, карта города*

*машина* → *стиральная машина, швейная машина*

# Близость запросов

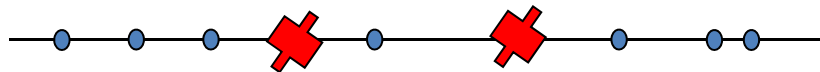


# Метод

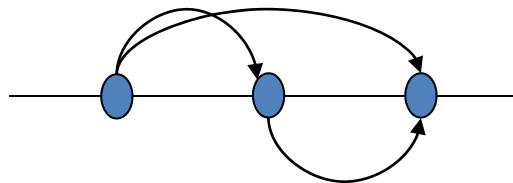
запрос=<timeStamp, userId, queryText, clicks>

0. чистка лога (порно, Яндекс, «подсказки», язык запросов, опечатки)

1. выделение сессий (по времени)



2. выделение пар (в.ч. транзитивно)



# Метод – 2

3. нормализация запросов (стоп-слова, капитализация, лемматизация, сортировка слов + выбор лучшего обратного преобразования)
4. борьба со флешмобами и «событийными» ассоциациями
5. матрица частоты переходов «запрос-запрос» (пороги для пользователя, ограничение на абсолютную частоту)
6.  $\text{weight}(q1 \rightarrow q2) = f(\text{freq}_{q1}, \text{freq}_{q2}, \text{freq}_{q1q2}) * f(\text{freq}_{q2q1})$
7. ранжирование, отсечение по порогу
8. индекс:  $q \rightarrow q1, q2, q3...$  (оригинальные запросы)

# Оценка

- «оценка глазами»
- «классы объектов»
- кластеры запросов Яндекс.Директа
- кликабельность



# Результат

«манго» — связанные запросы

[mango](#)   [манго одежда](#)   [авокадо](#)  
[папайя](#)   [манго фрукт](#)   [инжир](#)  
[финик](#)   [магазин манго](#)   [помело](#)

«умные фразы» — связанные запросы

[умные мысли](#)   [умные слова](#)   [смешные фразы](#)  
[умные цитаты](#)   [умные статусы](#)   [фразы о жизни](#)  
[умные вопросы](#)   [фразы о любви](#)   [мудрые фразы](#)

«хочу похудеть» — связанные запросы

[хочу любви](#)   [хочу умереть](#)   [хочу замуж](#)  
[хочу ребенка](#)   [хочу денег](#)

«смайлики» — связанные запросы

[смайлы для qip](#)   [анимашки](#)   [smiles](#)  
[улыбка](#)   [анимации](#)   [иконки](#)  
[позитив](#)

«люблю» — связанные запросы

[любимому смс](#)   [скучаю](#)   [сердце](#)  
[поцелуй](#)

«утконос» — связанные запросы

[ехидна](#)   [коала](#)   [кенгуру](#)  
[выхухоль](#)   [эму](#)   [муравьед](#)  
[вомбат](#)   [динго](#)   [броненосец](#)

«лук» — связанные запросы

[рыцарь вампир](#)   [лукашенко](#)   [мечи](#)  
[капуста](#)   [морковь](#)   [свекла](#)  
[огурец](#)   [картофель](#)   [арбалет](#)

«ива» — связанные запросы

[дуб](#)   [тополь](#)   [береза](#)  
[липа](#)   [сосна](#)   [клен](#)  
[ель](#)   [каштан](#)   [рябина](#)

# Литература

- Broder A. A Taxonomy of Web Search. SIGIR 2002.
- Broder A. et al. Robust classification of rare queries using web knowledge. SIGIR 2007.
- Risvik K. M. et al. Query Segmentation for Web Search. WWW2003.
- Bergsma S. & Wang Q. I. Learning Noun Phrase Query Segmentation. EMNLP-CoNLL 2007.
- Hagen M. et al. The Power of Naïve Query Segmentation. SIGIR 2010.
- Beitzel, S.M. et al. Temporal analysis of a very large topically categorized web query log, JASIST, vol. 58, no. 2, 2007.
- Beitzel, S.M., et al. Automatic classification of web queries using very large unlabeled query logs. ACM Trans. Inf. Syst., 25(2):9, 2007.
- Shen D. et al. Q2C@UST: Our Winning Solution to Query Classification in KDDCUP 2005, SIGKDD Explorations 7(2).
- Li X. et al. Learning Query Intent from Regularized Click Graphs. SIGIR 2008.

# Ссылки

- История про лог AOL – см. [http://en.wikipedia.org/wiki/AOL\\_search\\_data\\_scandal](http://en.wikipedia.org/wiki/AOL_search_data_scandal)
- Поиск по логу AOL2006 <http://www.aolstalker.com/>
- Query Log Analysis Workshop @ WWW2007, <http://querylogs2007.webir.org/>
- KDD Cup 2005, <http://www.sigkdd.org/kdd2005/kddcup.html>
- WSCD09: Workshop on Web Search Click Data 2009, <http://research.microsoft.com/users/nickcr/wscd09/>
- Microsoft Web N-gram Services, <http://research.microsoft.com/en-us/collaboration/focus/cs/web-ngram.aspx>
- Jiang D. et al. Web Search/Browse Log Mining: Challenges, Methods, and Applications, [http://research.microsoft.com/en-us/people/djiang/web\\_search\\_and\\_browse\\_log\\_mining.pdf](http://research.microsoft.com/en-us/people/djiang/web_search_and_browse_log_mining.pdf)

# ПРАКТИКА

# Практика

- Задачи
  - Сегментация
  - Тематическая классификация
  - Близкие запросы
- Данные – AOL2006
  - <http://www.gregsadetsky.com/aol-data/>
  - seed ~2000 запросов,  
[http://www.kansas.ru/querylog\\_analysis/2000queries.txt](http://www.kansas.ru/querylog_analysis/2000queries.txt)
  - + описания DMOZ для категорий второго уровня  
[http://narod.ru/disk/371552001/dmoz\\_data.zip.html](http://narod.ru/disk/371552001/dmoz_data.zip.html)
- Можно использовать любые внешние ресурсы (не человеческие ;)

# DMOZ

**dmoz** open directory project In partnership with **Aol Search.**


[about dmoz](#) | [dmoz blog](#) | [suggest URL](#) | [help](#) | [link](#) | [editor login](#)

[advanced](#)

<a href="#">Arts</a> <a href="#">Movies</a> , <a href="#">Television</a> , <a href="#">Music</a> ...	<a href="#">Business</a> <a href="#">Jobs</a> , <a href="#">Real Estate</a> , <a href="#">Investing</a> ...	<a href="#">Computers</a> <a href="#">Internet</a> , <a href="#">Software</a> , <a href="#">Hardware</a> ...
<a href="#">Games</a> <a href="#">Video Games</a> , <a href="#">RPGs</a> , <a href="#">Gambling</a> ...	<a href="#">Health</a> <a href="#">Fitness</a> , <a href="#">Medicine</a> , <a href="#">Alternative</a> ...	<a href="#">Home</a> <a href="#">Family</a> , <a href="#">Consumers</a> , <a href="#">Cooking</a> ...
<a href="#">Kids and Teens</a> <a href="#">Arts</a> , <a href="#">School Time</a> , <a href="#">Teen Life</a> ...	<a href="#">News</a> <a href="#">Media</a> , <a href="#">Newspapers</a> , <a href="#">Weather</a> ...	<a href="#">Recreation</a> <a href="#">Travel</a> , <a href="#">Food</a> , <a href="#">Outdoors</a> , <a href="#">Humor</a> ...
<a href="#">Reference</a> <a href="#">Maps</a> , <a href="#">Education</a> , <a href="#">Libraries</a> ...	<a href="#">Regional</a> <a href="#">US</a> , <a href="#">Canada</a> , <a href="#">UK</a> , <a href="#">Europe</a> ...	<a href="#">Science</a> <a href="#">Biology</a> , <a href="#">Psychology</a> , <a href="#">Physics</a> ...
<a href="#">Shopping</a> <a href="#">Clothing</a> , <a href="#">Food</a> , <a href="#">Gifts</a> ...	<a href="#">Society</a> <a href="#">People</a> , <a href="#">Religion</a> , <a href="#">Issues</a> ...	<a href="#">Sports</a> <a href="#">Baseball</a> , <a href="#">Soccer</a> , <a href="#">Basketball</a> ...
<a href="#">World</a> <a href="#">Català</a> , <a href="#">Dansk</a> , <a href="#">Deutsch</a> , <a href="#">Español</a> , <a href="#">Français</a> , <a href="#">Italiano</a> , <a href="#">日本語</a> , <a href="#">Nederlands</a> , <a href="#">Polski</a> , <a href="#">Русский</a> , <a href="#">Svenska</a> ...		

Help build the largest human-edited directory of the web

Copyright © 1998-2010 Netscape



категории второго уровня



4,771,467 sites - 89,784 editors - over 1,001,177 categories

Павел Браславский - Анализ запросов

# Примеры

## Сегментация

- at the vet | norman rockwell | litho
- big weenie | by eminem

## Тематическая классификация

- big weenie by eminem --> Arts\_Music /t Art\_Television

До трех категорий, упорядоченных по уменьшению уверенности

## Кластеризация

- 1996 mitsubishi mirage --> 2001 subaru impreza wagon /t toyota corolla /t toyota sienna hybrid

До 10 запросов из большого лога (сначала – самый близкий)

- Присылайте по почте с темой  
querylog\_analysis\_results
- segmentation\_имя\_фамилия.txt
- classification\_имя\_фамилия.txt
- clustering\_имя\_фамилия.txt
- имя\_фамилия.pdf - краткое описание  
методов



**Я**ндекс

**Павел Браславский**  
[pb@yandex-team.ru](mailto:pb@yandex-team.ru)