



ACM
SIGIR 2011

July 24-28, 2011

The 34th Annual International ACM SIGIR Conference



Web Retrieval: The Role of Users

Ricardo Baeza-Yates, Yahoo! Research, Barcelona, Spain

Yoelle Maarek, Yahoo! Research, Haifa, Israel



ABSTRACT

Web retrieval methods have evolved through three major steps in the last decade or so. They started from standard document-centric IR in the early days of the Web, then made a major step forward by leveraging the structure of the Web, using link analysis techniques. A more recent, no less important step forward, has been to enter the user in this equation in two ways: (1) **implicitly**, through the analysis of usage data captured by query logs, and session and click information in general, the goal being to improve ranking as well as to measure user's happiness and engagement; (2) **explicitly**, by offering novel interactive features; the goal here being to better answer users' needs. We cover here the user-related challenges associated with the implicit and explicit uses activities, namely:

Usage data analysis and metrics: It is critical to monitor how users interact with Web retrieval systems, as this implicit relevant feedback aggregated at a large scale can approximate quite accurately the level of success of a given feature. This is the focus of the first part of the tutorial.

User interaction: Web retrieval engines interact with the user at two key stages, each associated with its own challenges: expressing a query and interpreting and using results. After quite some stagnation on the front-end of Web retrieval, we have seen numerous novel interactive features appear in the last 3 to 4 years, as the leading commercial search engines seem to compete for users' attention. The second part of the tutorial will be dedicated to explicit user interaction. We will introduce novel material (as compared to previous versions of this tutorial.)

An earlier version of this tutorial was given at
SIGIR'2010, WSDM'2011 and ECIR'2011



Table of contents

- Session 1: Introduction2-7
 - Web search - retrospective
 - Major game changers
- Session 2: Implicit user aspects8-59
 - Mining query logs
 - Monitoring user's behavior (click models)
- Session 3: Explicit user aspects60-96
 - Before the query is issued
 - Search results, traditional, interactive and actionable
 - Query assistance
 - Query introspection
- Session 4: Conclusions.....101-104
 - Current and next challenges
- Bibliography



Looking back

1970-90

Information Retrieval

1990-96

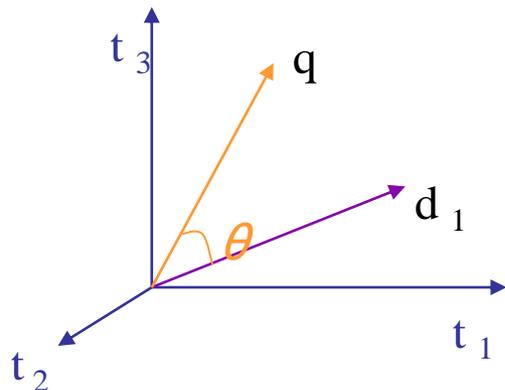
Analyze words in millions of Web pages

Compare queries and documents as vectors of words

– Information Retrieval “invented” by Gerry Salton at Cornell: Text analysis, count words [Salton’86] “Introduction to Modern Information Retrieval”

– Vector Space model (d_1 document, q query)

- Evaluation on predefined test collections (e.g., 300 docs CACM)



Excite, Lycos, Inktomi, Alta-Vista

- HTTP (CERN) and the Web,
- First browser: Mosaic (U of Illinois)
- First crawler by Fuzzy Mauldin (Lycos’ founder)
- Yahoo! founded as a directory service,
- Evaluation by TREC (NIST)



- **Classic IR**

- **Corpus:** Fixed document collection
- **Goal:** Retrieve documents with information content that is relevant to user's information need

- **Classic relevance**

- For each query Q and stored document D in a given corpus assume there exists a relevance score $\text{Score}(Q, D)$, where

$\text{Score}(Q, D)$ is averaged over users U and contexts C

- Optimize $\text{Score}(Q, D)$ as opposed to $\text{Score}(Q, D, U, C)$
- In other words, usually:

1. Corpus is **predetermined**
2. Context is **ignored**
3. Users are **ignored**

Bad assumptions
in the web context



Then first revolution

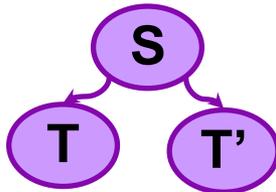
1998: Analyze links on a very large scale

Google PageRank

- o [Brin & Page WWW'1998], "The anatomy of a large-scale hypertextual web search engine",

IBM Hubs and Authorities

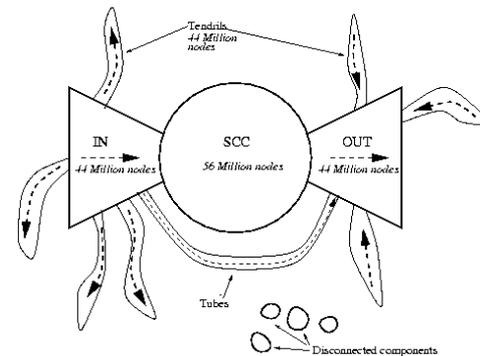
- o [Kleinberg SIAM'1998], "Authoritative sources in a hyperlinked environment"



- o Recommendation of T by S
- o Evidence that S and T are related
- o Relation between T and T'

Structure and connectivity

- o [Broder et al. WWW 2000] "Graph structure of the Web"



Anchor text is a key signal

- o [Eiron & McCurley SIGIR'2003] "Analysis of anchor text for web search"



Then the second revolution

2000 Usage Data

AOL query log transcript

AnonID	Query	QueryTime	ItemRank	ClickURL
100218	tennessee department of transportation	2006-03-01 11:08:30	1	http://www.tdot.state.tn.us
100218	tennessee federal court	2006-03-01 11:53:44	1	http://www.constructionweblinks.com
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	2	http://www.tennessee.gov
100218	state of tennessee emergency communications board	2006-03-01 12:56:18	1	http://www.tennessee.gov
100218	dixie youth softball	2006-03-02 10:36:48	2	http://www.dixie.org
100218	cdwg	2006-03-03 14:29:07	1	http://www.cdwg.com
100218	cdwg scam cdwge	2006-03-03 14:30:11	1	http://www.cdwg.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	2	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	escambia county sheriff's department	2006-03-07 09:26:51	1	http://www.escambiaso.com
100218	pensacola police department	2006-03-07 09:34:28	1	http://www.pensacolapolice.com
100218	memphis pd	2006-03-07 09:42:33	1	http://www.memphispolice.org
100218	nashville metro pd	2006-03-07 09:44:43	1	http://www.police.nashville.org
100218	florida highway patrol	2006-03-07 09:48:35	1	http://www.fhp.state.fl.us
100218	tennessee highway patrol	2006-03-07 09:49:52	1	http://www.state.tn.us
100218	florida bureau of investigations	2006-03-07 09:51:08	2	http://www.flbvi.com
100218	florida bureau of investigations	2006-03-07 09:51:08	1	http://www.flbvi.com
100218	government finance officers association	2006-03-07 21:16:11	1	http://www.massdor.com
100218	state of tennessee controllers manual	2006-03-07 21:17:12	1	http://www.massdor.com
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	3	http://www.comptroller.state.tn.us
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	4	http://www.fbr.state.tn.us
100218	state of tennessee audit controllers manual	2006-03-07 21:17:40	9	http://audit.tennessee.edu
100218	internal controls for municipalities under 10 000	2006-03-07 21:38:04	1	http://www.myscpa.org
100218	internal controls for municipalities under 10 000	2006-03-07 21:38:04	4	http://www.massdor.com
100218	municipality fraud detection techniques	2006-03-07 21:41:40	1	http://www.massdor.com
100218	municipal fraud audit detection internal controls	2006-03-07 21:43:15	1	http://www.massdor.com
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	1	http://www.whitehouse.gov
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	4	http://www.hhgc.org
100218	internal fraud controls for municipalities cities towns local government	2006-03-07 21:45:13	7	http://www.sao.state.ut.us
100218	evaluating internal controls a local government managers guide	2006-03-07 21:51:18	5	http://www.allbusiness.com

Implicit feedback

Query logs and Click through data

Live experiments

Try out experimental features on % of users



[Xiaodong Shi, Technical report, University of Michigan, 2007]
"Social Network Analysis of Web Search Engine Query Logs"

[Andy Beal, Market Pilgrim 2005]

"Google Space project at London's Heathrow airport and lab mice in a maze?"





Today: users rule

- **2000-to-date: Usage data & live experiments**
 - A wealth of **implicit** information
 - Query logs and CTR represent “implicit relevant feedback”
 - [Srikant et al. KDD'2010] “*User Browsing Models: Relevance versus Examination*”
 - Increased traffic generates more and more usage data
 - Wisdom of crowds works (see “did you mean”)
 - Dashboards and 1% AKA bucket experiments
 - Users’ actions give competitive advantage to the engine ranking
 - **Explicit** information from more and more sophisticated users
 - Front-end apps and features are the key differentiator
 - Launch and iterate!



Goals of today's tutorial

- Define and organize concepts you are all familiar with, as users.
- Identify the technical/research challenges
- Some hints on how it works
- Some understanding on the explicit/implicit role of users in Web search

Disclaimer!!!

Most (not all) of this lecture is high-level as most algorithms and implementation details are still kept as trade secrets by most search engines



Implicit users aspects

Session 2



Collective usage data

- Usage data is the new entry barrier for any search engine
- The more it grows, and the better mining technologies get, the more relevant search results become
- Collective usage data is key to
 1. Interpreting users' information needs
 2. Improving ranking and all search artifacts



Query logs: mining queries for ...

Improving Web search

- Index layout and caching
[Baeza-Yates et al, SIGIR 2007] “The Impact of Caching on Search Engines”
- Features for ranking
- Query intention
- Query recommendations
- Matching advertising



Mining queries for ...

- User search modelling
 - [Baeza-Yates, Hurtado, Mendoza & Dupret, LA-WEB'2005] "Modeling User Search Behavior"
- Discoverability for search engines & people
- User driven Web design
 - The Web site that the users want
 - Improve information scent, content & structure
- Bootstrap pseudo-semantic resources



- Cultural and educational diversity
- Short queries & impatient interaction
 - few queries posed & few answers seen
- Smaller & different vocabulary
- Different user goals

[Broder, SIGIR Forum 2002] "A Taxonomy of Web Search"

- Information needs
- Navigational needs
- Transactional needs

See refinement by [Rose & Levinson, WWW 2004] "Understanding user goals in web search"



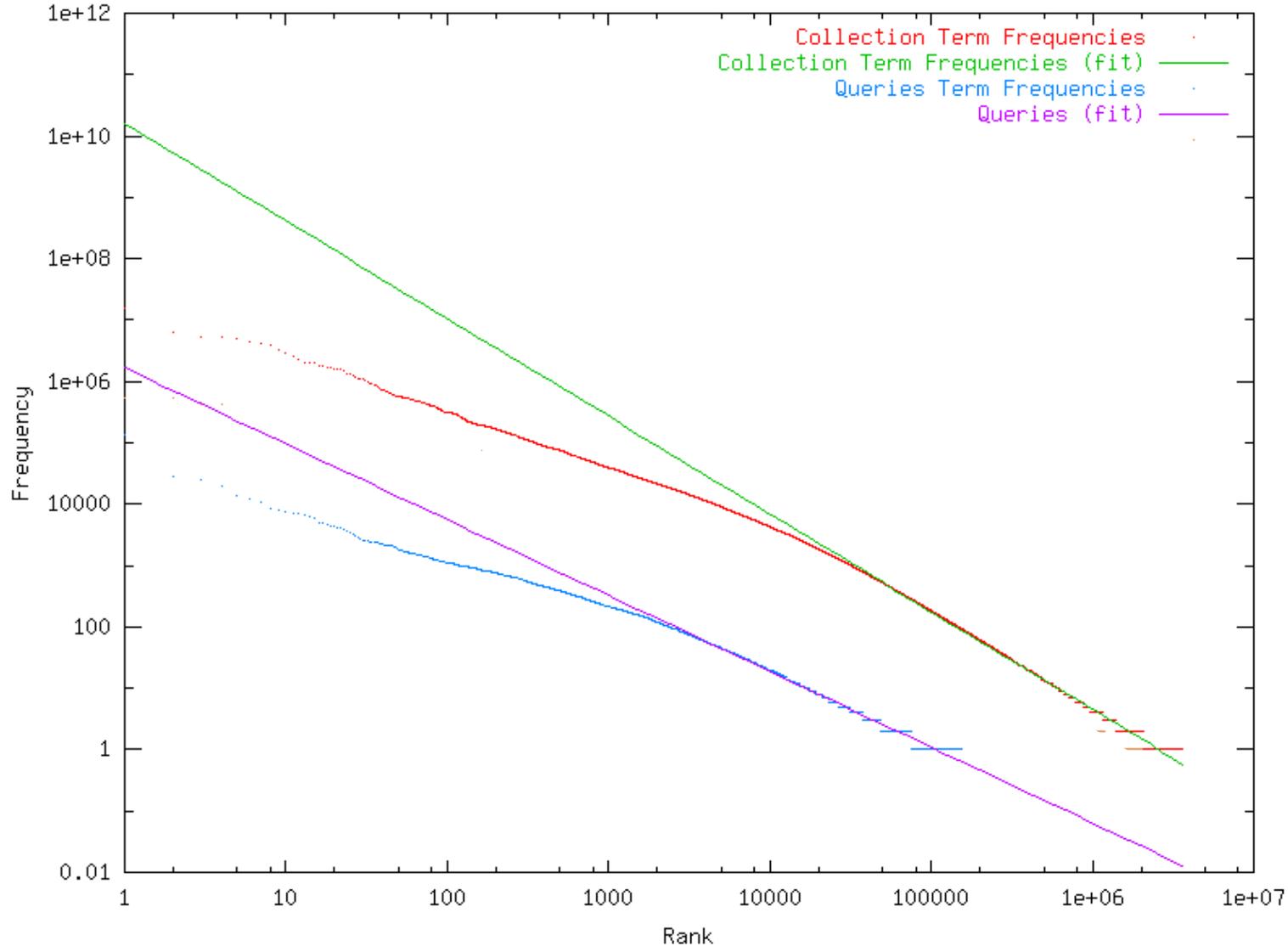
Who issues head/tail queries?

- Long tail phenomenon
 - Popular hypothesis: Majority of consumers consistently follow the crowd, only eccentrics issue tail queries
- Study of long tail
[Goel et al. WSDM'2010] "*Anatomy of the Long Tail: Ordinary People with Extraordinary Tastes*"
 - Extensive study on user preferences for movies, music, Web browsing and **Web search**
 - Everyone is a bit eccentric, consuming both popular and specialty products, e.g., most people have their own 80-20.
 - Supporting the tail (products or queries) goes beyond direct revenues to second-order gains associated with
 - increased consumer satisfaction
 - repeat patronage.
 - Supporting the tail may boost the head by providing users a one-stop shop for both their mainstream and niche interests.



Queries and text

Word distribution in queries and in documents are different

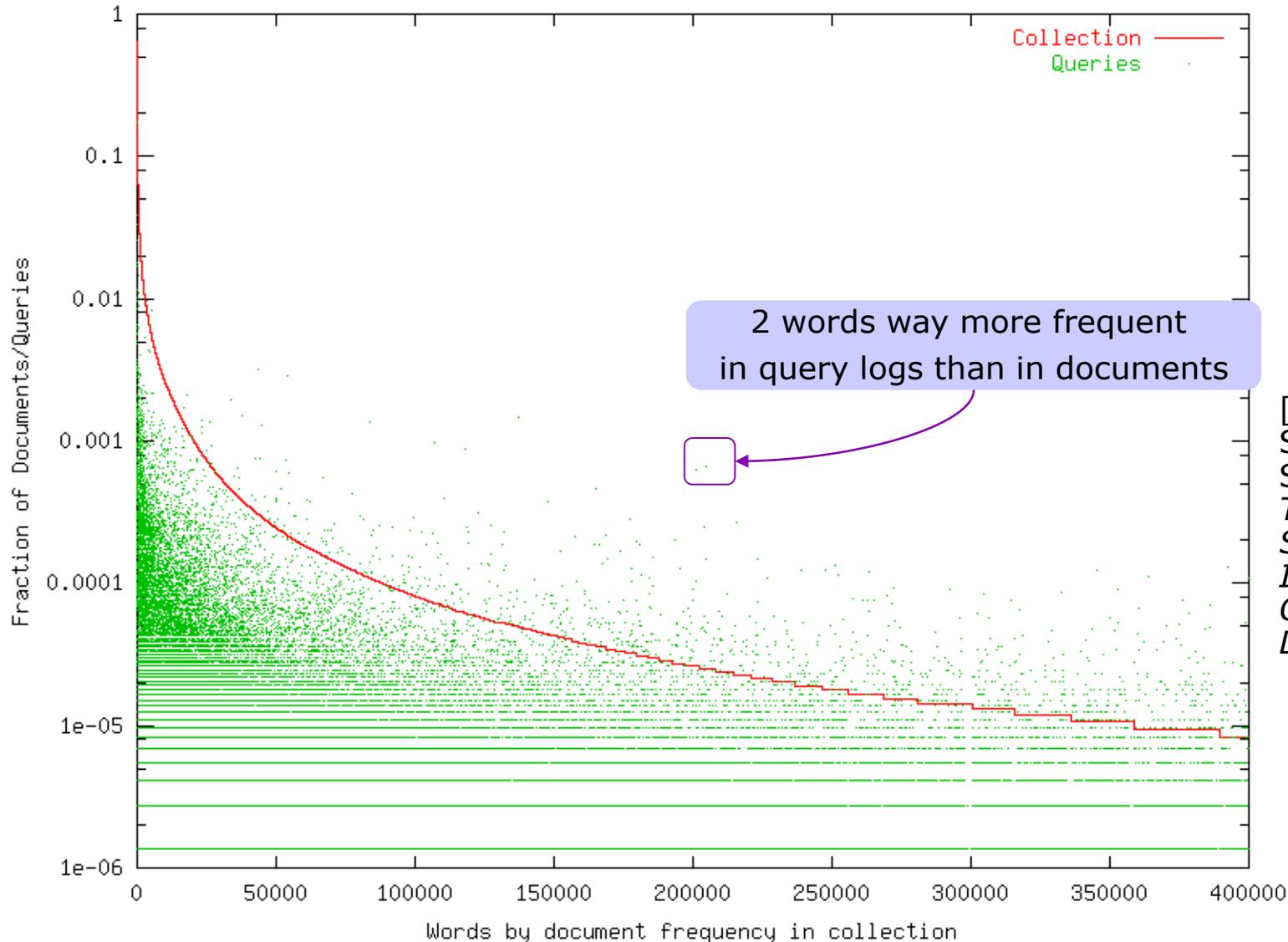


[Baeza-Yates & Saint-Jean, SPIRE'2003] "A Three Level Search Engine Index based in Query Log Distribution"



Queries and text

Word distribution in queries in documents now sorted such that same word on x



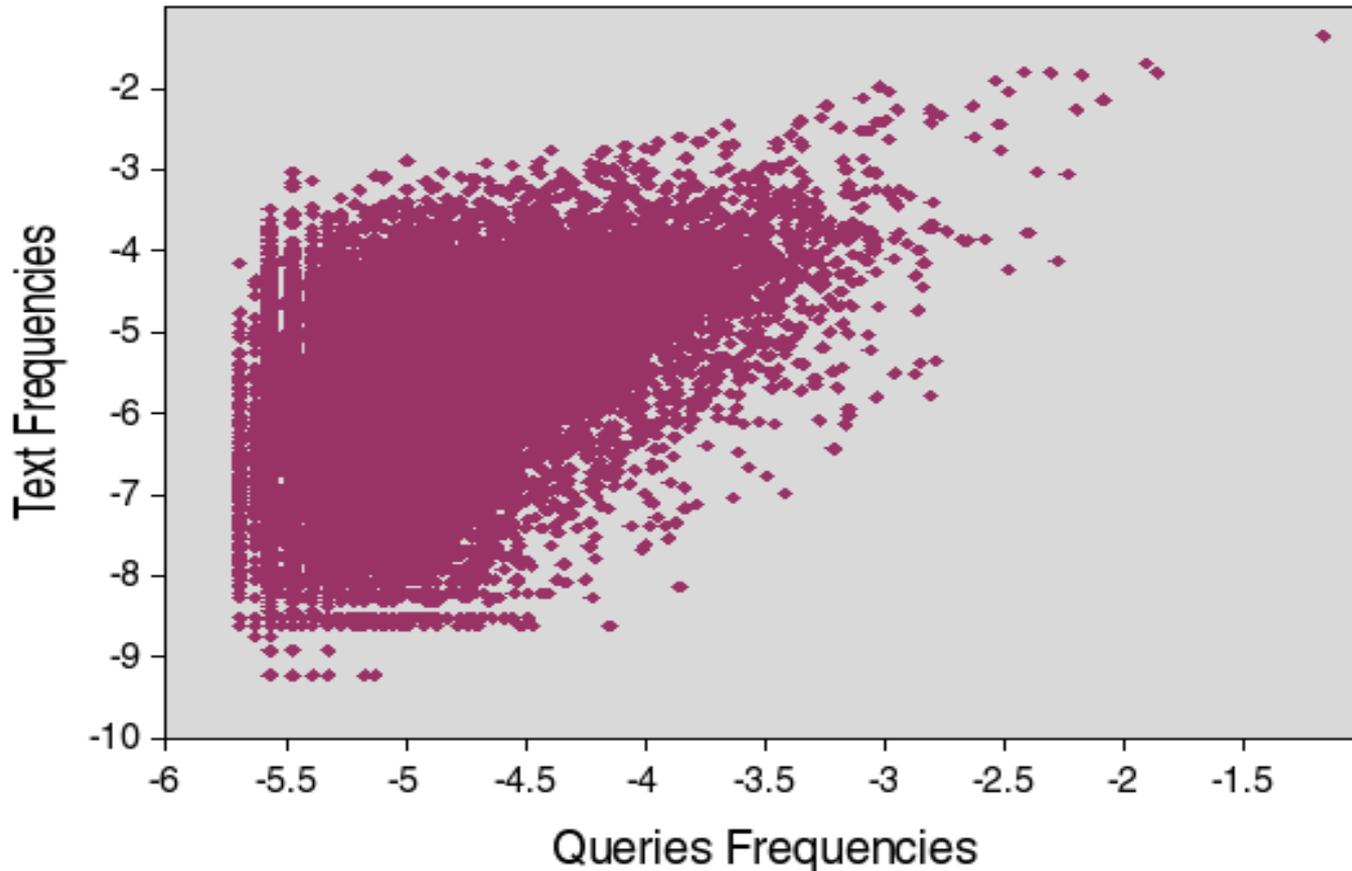
[Baeza-Yates & Saint-Jean, SPIRE'2003] "A Three Level Search Engine Index based in Query Log Distribution"



Queries and text

Log-log plot – weak correlation
bw query & document vocabularies

Term Pairs

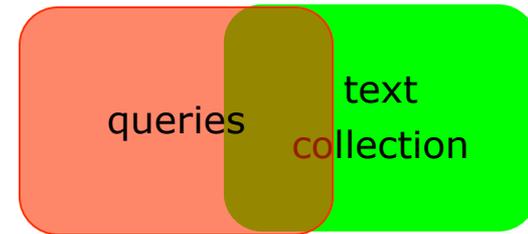
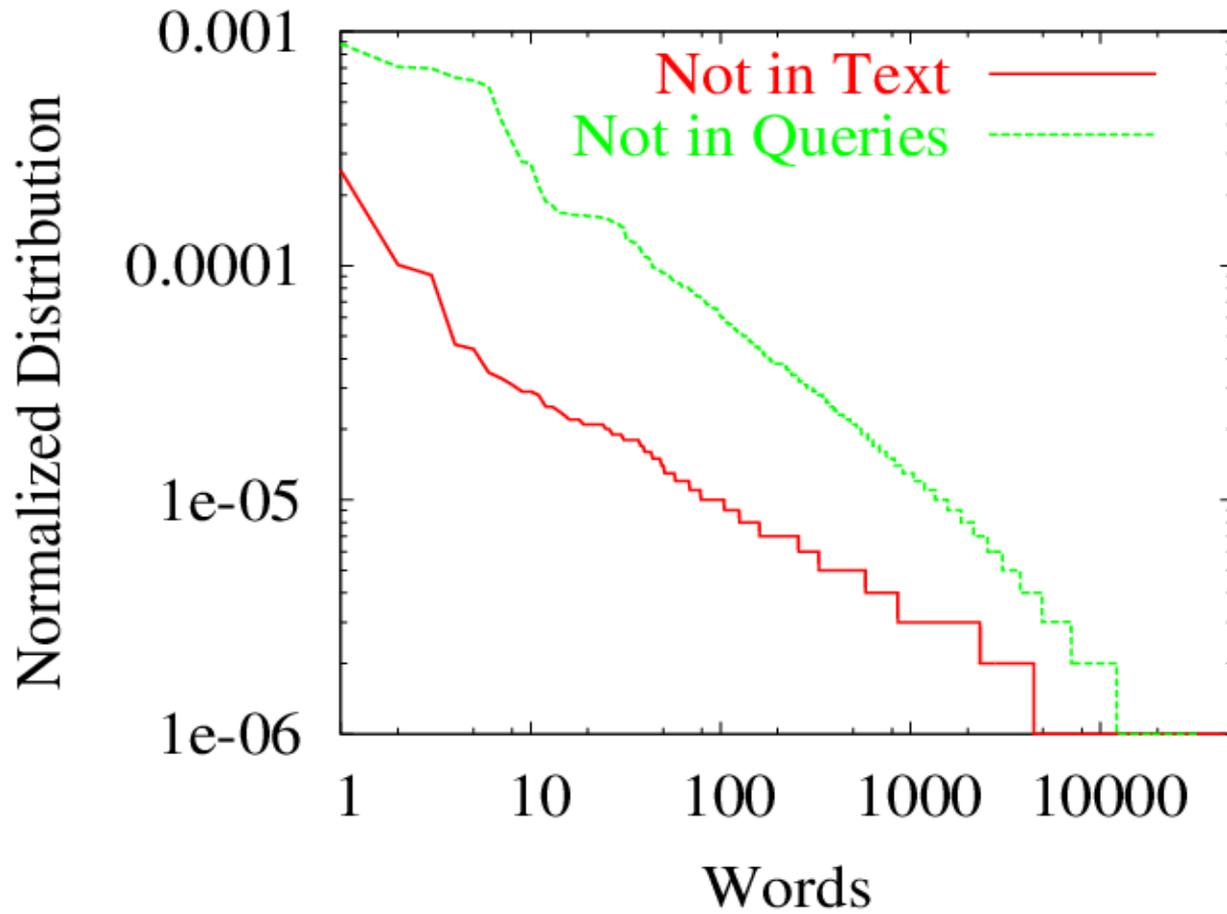


[Baeza-Yates et al, SIGIR 2007]
"The Impact of Caching on Search Engines"



Queries and text

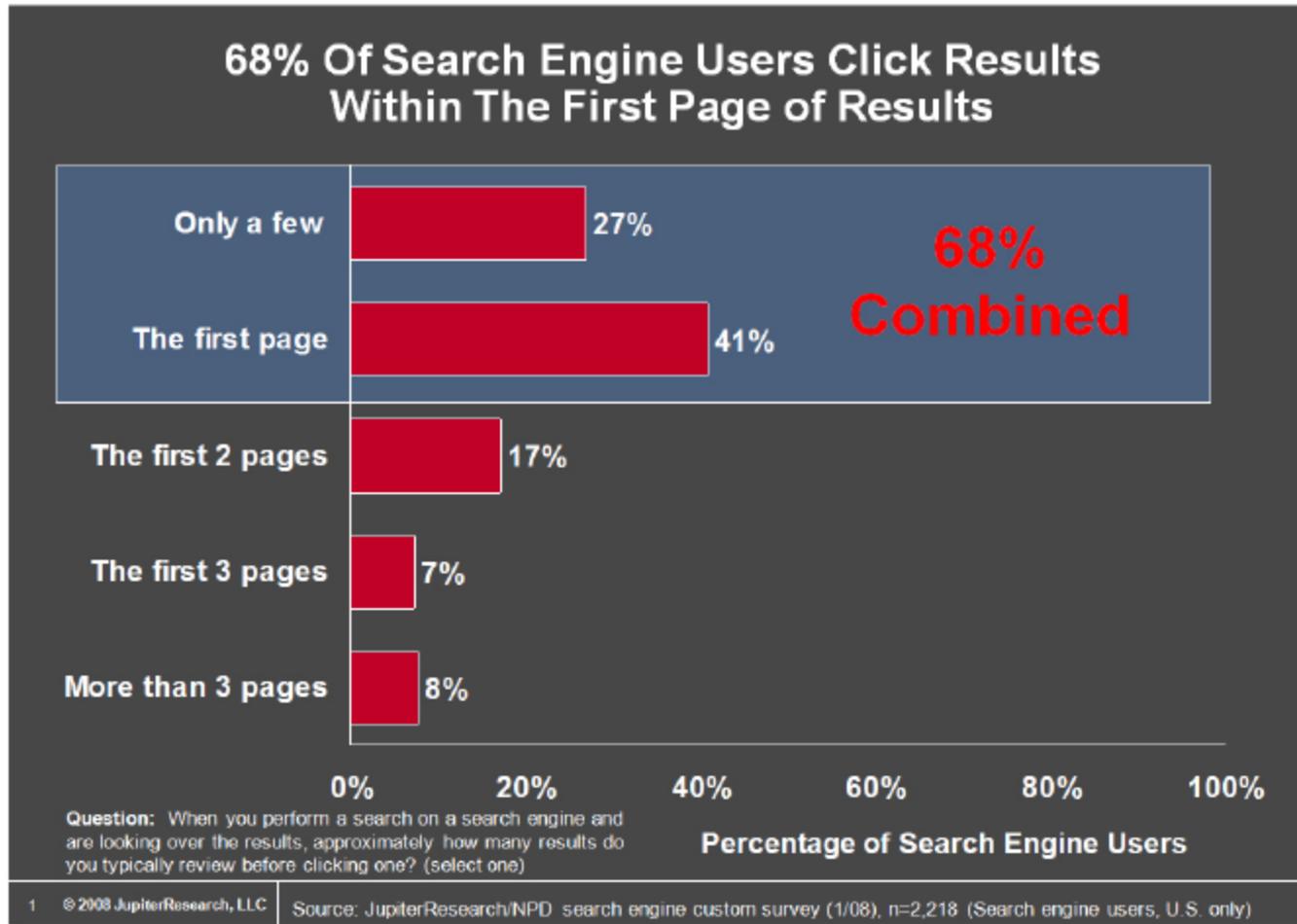
Each set with its own power law!





Other implicit signals

How far do people look for results?



Source: iprospect.com *iProspect Blended Search Results Study – April 2008*



Evolution of behavior

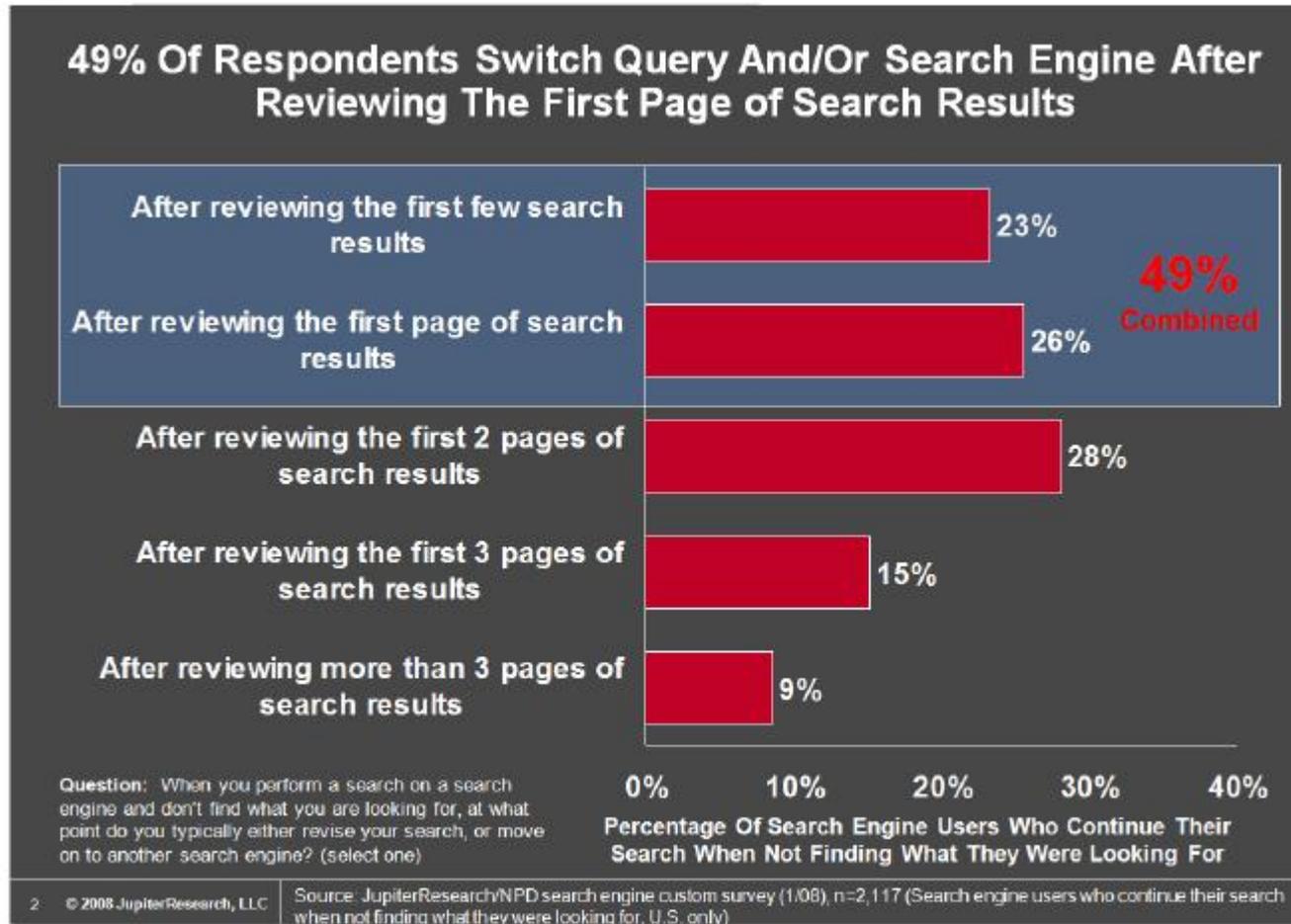
“The first three pages of search results now appear to be the *last frontier*”

	2008	2006	2004	2002
Only a few	27%	23%	24%	16%
The first page	41%	39%	36%	32%
The first 2 pages	17%	19%	20%	23%
The first 3 pages	7%	9%	8%	10%
More than 3 pages	8%	10%	12%	19%

Source: iprospect.com *iProspect Blended Search Results Study – April 2008*



“When you [...] don’t find what you are looking for, [when] do you [...] revise your search, or move on to another search engine?”



Source: iprospect.com *iProspect Blended Search Results Study – April 2008*





Same question over time

	2008	2006	2004	2002
After reviewing the first few search results	23%	16%	23%	14%
After reviewing the first page of search results	26%	25%	19%	14%
After reviewing the first 2 pages of search results	28%	27%	26%	28%
After reviewing the first 3 pages of search results	15%	20%	15%	22%
After reviewing more than 3 pages of search results	9%	12%	17%	22%

Source: iprospect.com *iProspect Blended Search Results Study – April 2008*



Counting queries

Simplest method: count! and compare

Application: spelling – from <http://www.google.com/jobs/britney.html>

488941	brjtney spears	29	britent spears	9	brinttany spears	5	brney spears	3	britiy spears	2	brirreny
40134	brittany spears	29	brittnany spears	9	britanay spears	5	broitney spears	3	britmeny spears	2	brittany
36315	brittney spears	29	britttany spears	9	britinany spears	5	brotny spears	3	britneey spears	2	brirttany
24342	britany spears	29	btiney spears	9	brin spears	5	bruteny spears	3	britnehy spears	2	brirttney
7331	britny spears	26	birttney spears	9	britnew spears	5	btiyney spears	3	britnely spears	2	britainy
6633	briteny spears	26	breitney spears	9	britneyn spears	5	btrittney spears	3	britnesy spears	2	britaney
2696	britteny spears	26	brinity spears	9	britrney spears	5	gritney spears	3	britnetty spears	2	britaney
1807	briney spears	26	britenay spears	9	brtiny spears	5	spritney spears	3	britnex spears	2	britanay
1635	brittny spears	26	britneyt spears	9	brtittney spears	4	bitny spears	3	britneyxxx spears	2	britanay
1479	brintey spears	26	brittan spears	9	brtny spears	4	bnritney spears	3	britnity spears	2	britanay
1479	britanny spears	26	brittne spears	9	brytny spears	4	brandy spears	3	britney spears	2	britanay
1338	britiny spears	26	btittany spears	9	rbitney spears	4	brbritney spears	3	britnyey spears	2	britanay
1211	britnet spears	24	beitney spears	8	birtiny spears	4	breatiny spears	3	britterny spears	2	britanay
1096	britiney spears	24	birteny spears	8	birtney spears	4	breetney spears	3	brittney spears	2	britanay
991	britaney spears	24	brightney spears	8	brattany spears	4	bretiney spears	3	brittney spears	2	britanay
991	britnay spears	24	brintiny spears	8	breitny spears	4	brfitney spears	3	brittnyey spears	2	briteen
811	brithney spears	24	britanty spears	8	breteny spears	4	briattany spears	3	brityen spears	2	britenay
811	brtiney spears	24	britenny spears	8	brightny spears	4	brieteny spears	3	briytney spears	2	britenet
664	birtney spears	24	britini spears	8	brintay spears	4	briety spears	3	brltney spears	2	britenay
664	brintney spears	24	britnwy spears	8	brinttey spears	4	brittny spears	3	broteny spears	2	briteny
664	briteney spears	24	brittni spears	8	briotney spears	4	briittany spears	3	brtaney spears	2	britanay
601	bitney spears	24	brittnie spears	8	britanys spears	4	brinie spears	3	brtiiany spears	2	britinay
601	brinty spears	21	biritney spears	8	britley spears	4	brinteney spears	3	brtinay spears	2	britinay
544	brittaney spears	21	birtany spears	8	britneyb spears	4	brintne spears	3	brtinney spears	2	brityay
544	brittnay spears	21	biteny spears	8	britrney spears	4	britaby spears	3	brtitaney spears	2	britanay
364	britey spears	21	bratney spears	8	britrny spears	4	britaey spears	3	brtiteny spears	2	britrnat
364	brittiny spears	21	britani spears	8	brittner spears	4	britainey spears	3	brtnet spears	2	britrba
329	brtney spears	21	britanie spears	8	brottany spears	4	britinie spears	3	brytiny spears	2	britrnd
269	broney spears	21	bratany spears	8	britrney spears	4	britrney spears	3	brtney spears	2	britrn





Did you mean?

- Revolutionized spelling correction by departing from the usual dictionary-based model.
 - Classic approach was to use edit distances to identify typing mistakes such as letter inversions [Kukich, ACM Computing Surveys 1992] "*Techniques for automatically correcting words in text*"
 - Instead, "Did you mean" learns its spelling corrections simply from usage and a great deal of usage.
- Extensive use of query logs analysis
 - Frequency
 - Transition
 - Clicks

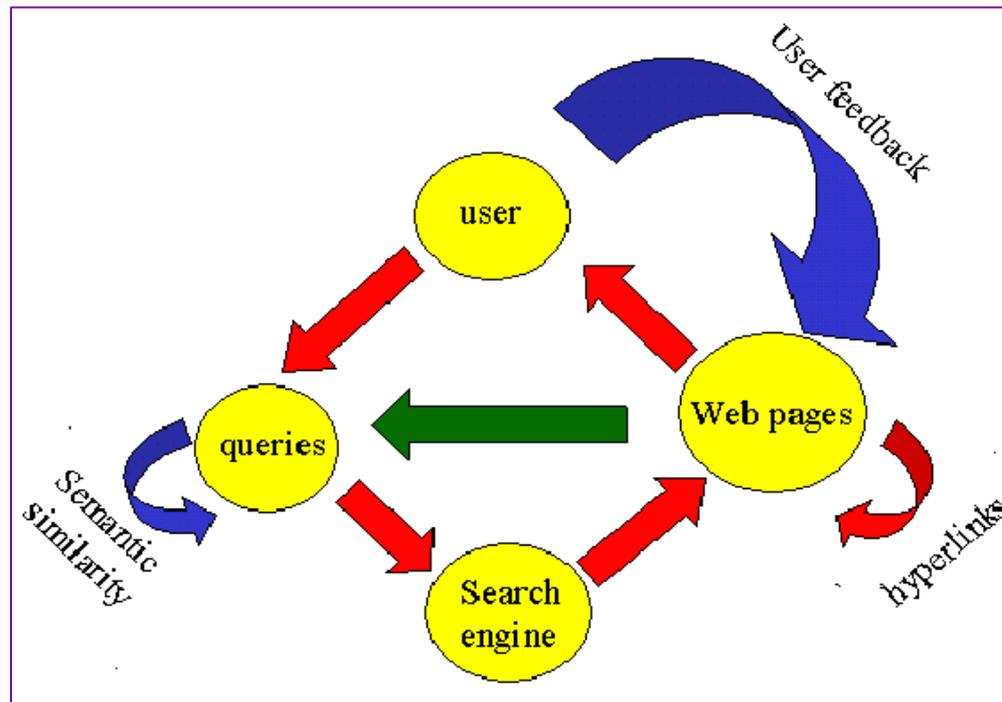


- Define relations among queries
 - Common words: sparse set
 - Common clicked URLs: better
 - Natural clusters
- Define distance function among queries
 - Content of clicked URLs [Baeza-Yates, Hurtado & Mendoza, Web Intelligence 2004] "*Ranking Boosting based in Query Clustering*"
- Using search results to measure the similarity between queries
 - [Sahami and Heilman WWW'2006] "*A Webbased Kernel Function for Measuring the Similarity of Short Text Snippets*" who propose a method for:
 - “measuring the similarity between short text snippets (even those without any overlapping terms) by leveraging web search results to provide greater context for the short texts”



Challenges

- Can we cluster queries well?
- Can we assign user goals to clusters?





Evaluating query similarity

- Cluster text of clicked pages
 - Infer query clusters using a vector model

$$q[i] = \sum_{URLu} \frac{\text{Pop}(q, u) \times \text{Tf}(t_i, u)}{\max_t \text{Tf}(t, u)}$$

- Unbias the effect of the rank and the interface in the clicks
- Pseudo-taxonomies for queries
 - Real language (slang?) of the Web
 - Can be used for classification purposes



Using the clusters

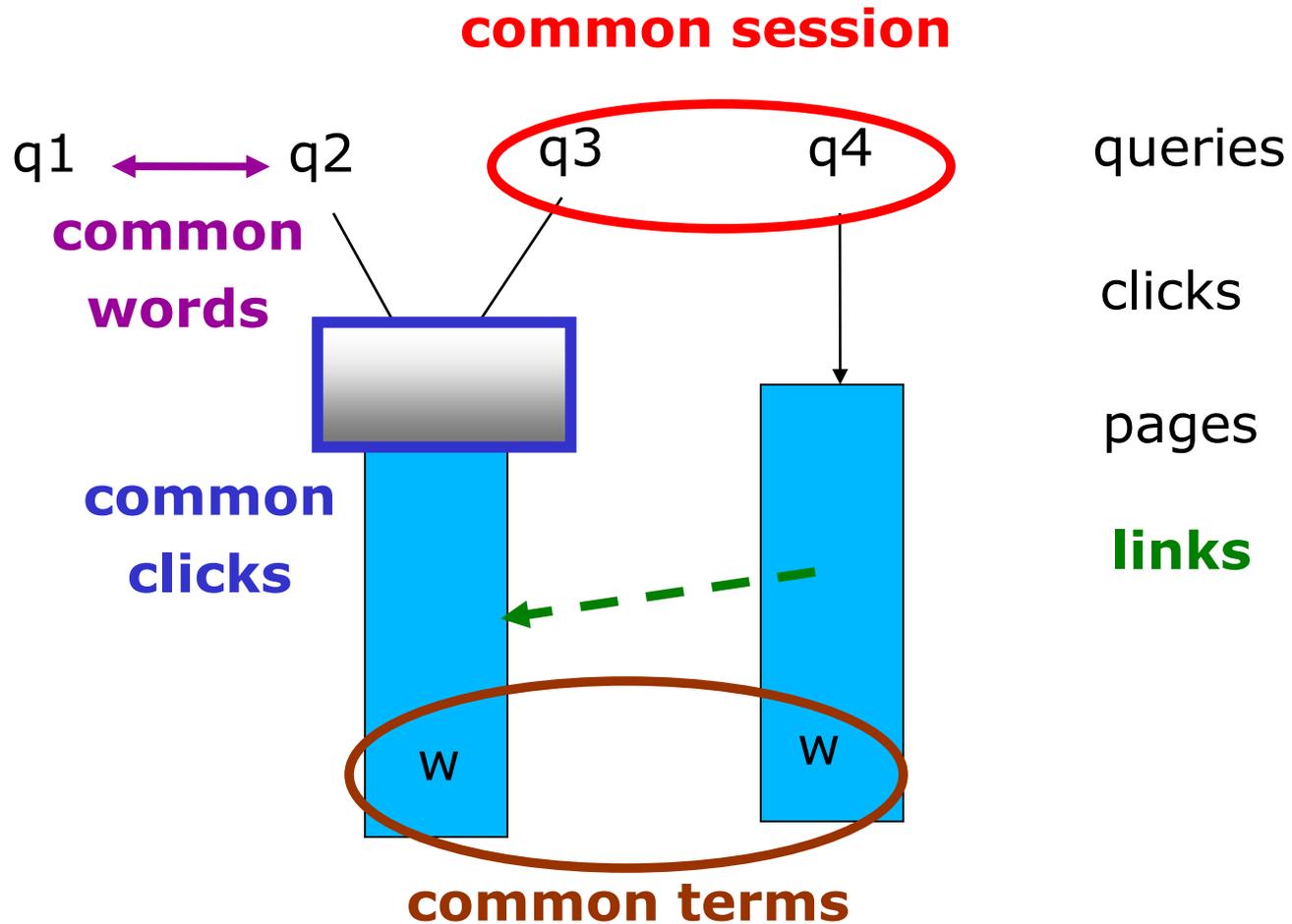
- Improved ranking
 - [Baeza-Yates, Hurtado & Mendoza, JASIST 2007] "*Improving search engines by query clustering*"
- Word classification
 - Synonyms & related terms are in the same cluster
 - Homonyms (polysemy) are in different clusters
- Query recommendation (ranking queries!)
 - Real queries, not query expansion

$$\text{Rank}(q) = \gamma \times \text{Sup}(q, q_{ini}) + (1 - \gamma) \times \text{Clos}(q, q_{ini})$$



Relating queries

[Baeza-Yates, SOFSEM'2007- Theory and Practice of Computer Science]
"Graphs from Search Engine Queries"



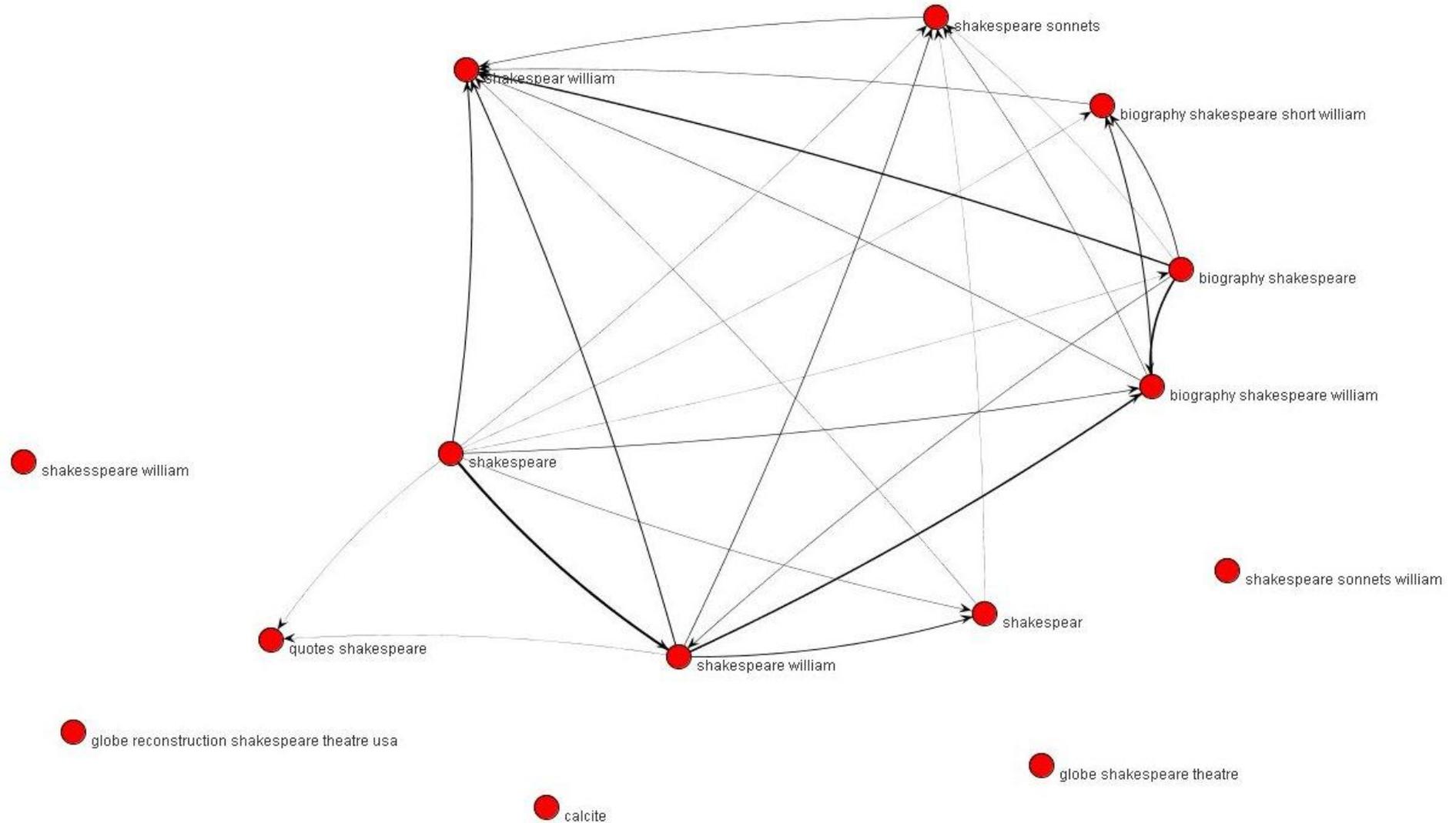


Qualitative analysis

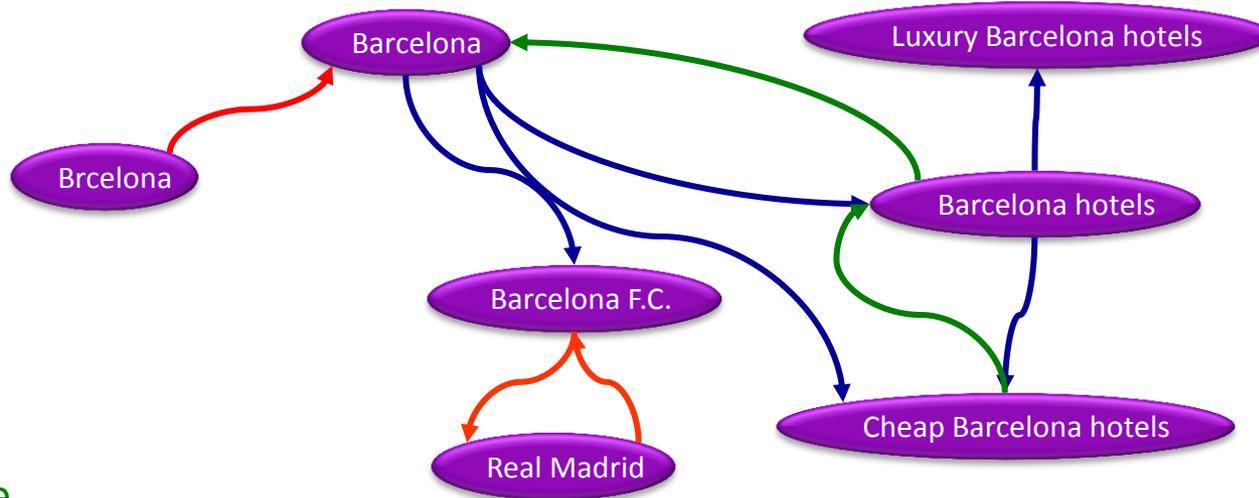
Graph	Strength	Sparsity	Noise
Word	Medium	High	Polysemy
Session	Medium	High	Physical sessions
Click	High	Medium	Click spam
Link	Weak	Medium	Link spam
Term	Medium	Low	Term spam



Query log: edge=common result click



Beyond counting: query flow analysis



- Correct
- Specialize
- Generalize
- Parallel Move

[Boldi, Bonchi, Castillo and Vigna Web Intelligence 2009] *“From Dango to Japanese Cakes: Query Reformulation Models and Patterns”*

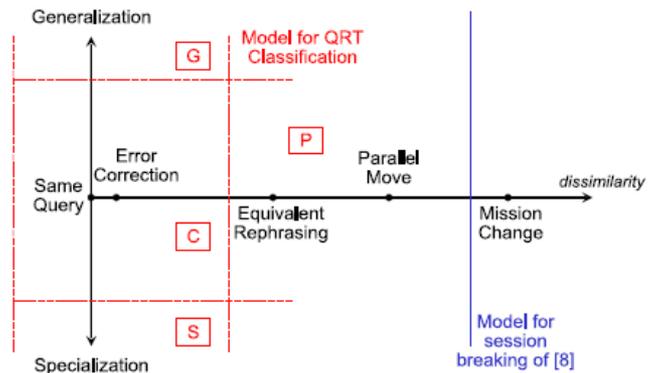


Figure 1. Graphical depiction of transition types.



Mining query logs + clicks for “research sessions”

[Donato, Bonchi, Chi & Maarek WWW 2010] “Do you want to take notes? Identifying research missions in Yahoo! Search Pad”

- **Research missions**
 - Defined as a “set of related and complex information needs”
 - Travel needs
 - Education needs
 - Medical needs
- **Research missions are conducted during research sessions**
 - Defined as a set of all “user activities (queries and clicks) occurring during a research mission”
 - May contained queries from distinct sessions
 - Key signals: topical coherence and user’s engagement
 - Account for
 - 10% of the search sessions
 - 25% of query volume



Research session - a formal definition

- A research session R is a maximal order sequence

$$R = \left\langle \left\langle q_{i_1}, u_{i_1}, t_{i_1}, C_{i_1} \right\rangle, \dots, \left\langle q_{i_k}, u_{i_k}, t_{i_k}, C_{i_k} \right\rangle \right\rangle$$

Where, for given thresholds s_θ , t_θ and c_θ , we have

$$u_{i_1} = \Lambda = u_{i_k} = u \in U \text{ and } t_{i_1} \leq \Lambda \leq t_{i_k} \leq \tau$$

$$\forall l, j \in \{i_1, \dots, i_k\}: s(f(q_l), f(q_j)) \geq s_g$$

$$|R| = k \geq k_g$$

$$\sum_{j=1}^k |C_{i_j}| \geq c_g$$

No specific time limitation

Topical coherence

Complexity

Engagement



- **Features extractor:** computes ~ 30 features from queries stream
 - Textual features:
 - size of the intersection on character-level 3-grams
 - cosine similarity computed on sets of stemmed words
 - Session features: number of clicks and queries since the beginning of the session
 - Time related features: interval between two queries
- **Research detector:** boosted decision tree that decides whether two queries q_1 and q_2 belong to the same research task
 - $q_1 = \text{house buying in Palo Alto}$ $q_2 = \text{Real estate Bay Area}$ $q_3 = \text{pizza hut}$
 - $\text{research}(q_1, q_2) = 1$ $\text{research}(q_2, q_3) = 0$
- **Mission boundaries detector:** boosted decision tree that decides if two queries q_1 and q_2 are topically coherent and thus susceptible to be part of the same mission.



An application: Search Pad* More details in Session 3

The screenshot shows a Yahoo! search results page for the query "barcelona airport transfers". The page includes the Yahoo! logo, navigation tabs (Web, Images, Video, Local, Shopping, More), and a search bar containing the query. Below the search bar, there is a "View Notes (9)" button and a "Take notes on barcelona cheap flights?" prompt with "Yes" and "No" buttons. The search results are displayed in a list format, with the first result being a sponsored advertisement for "Airport Transfer Barcelona" from Viator. Other results include "Barcelona Airport Transfers | Transportation from Barcelona ...", "Barcelona Airport Transfer", "Barcelona Airport Transport", and "Barcelona airport shuttle, airport shuttle transfer and ...".

Annotations on the screenshot include:

- A purple box highlights the top navigation area, including the Yahoo! logo and the search bar.
- A purple arrow points from the text "Appears when research session is identified" to the search bar area.
- A purple box highlights the "Take notes on barcelona cheap flights?" prompt.

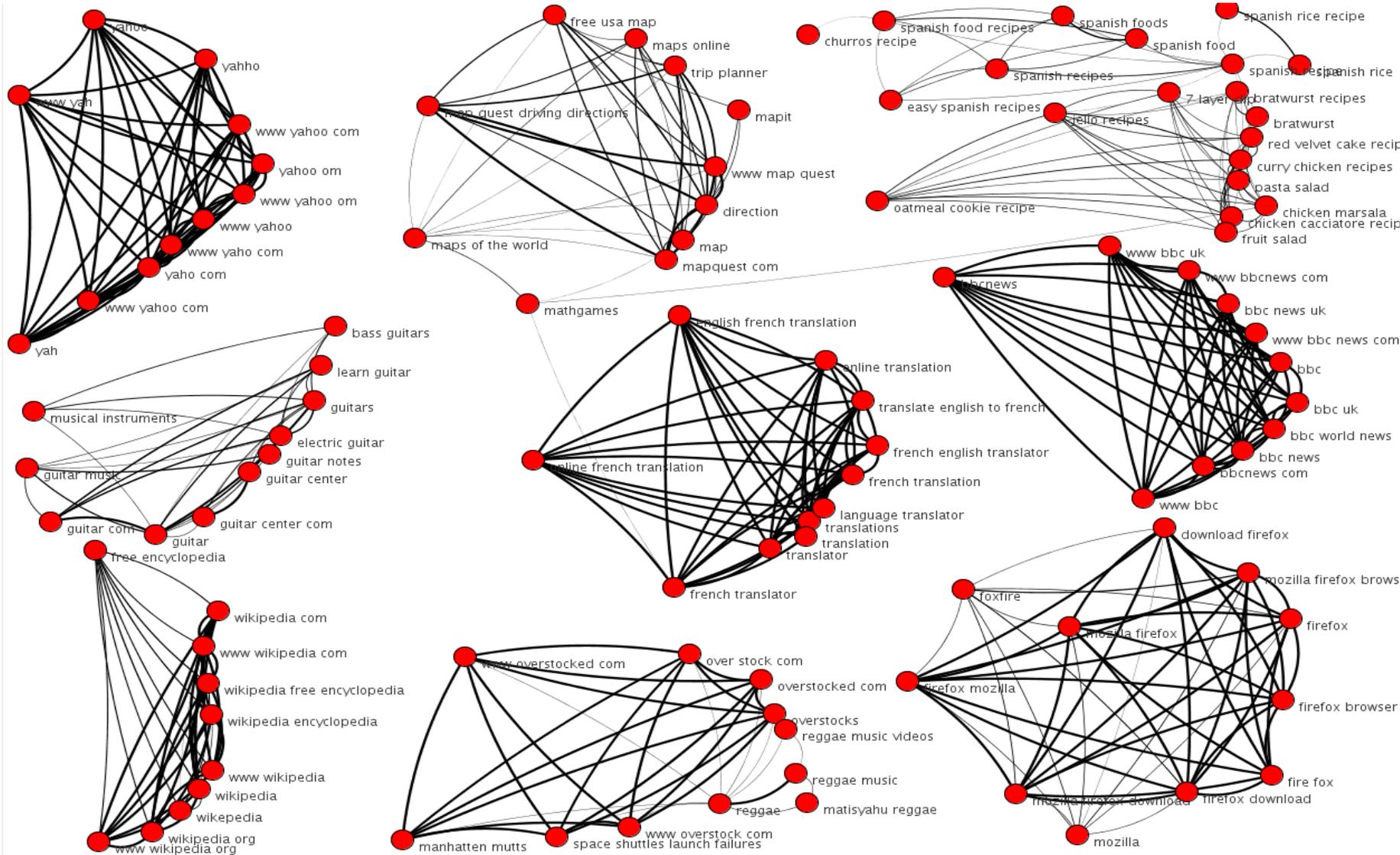
Appears when research session is identified

*removed on April 20- web search interface is a moving ground!





Beyond search: implicit folksonomy?





Set relations

- Identical sets: **equivalence**
- Subsets: **specificity**
 - directed edges
- Non empty intersections (with threshold)
 - degree of relation
- Dual graph: URLs related by queries
 - High degree: multi-topical URLs



Evaluation: ODP similarity

- A simple measure of similarity among queries using ODP categories
 - Define the similarity between two categories as the length of the longest shared path over the length of the longest path
 - Let c_1, \dots, c_k and c'_1, \dots, c'_k be the top k categories for two queries. Define the similarity ($@k$) between the two queries as

$$\max\{sim(c_i, c_j) \mid i, j = 1, \dots, k\}$$

[Baeza-Yates & Tiberi KDD' 2007] "*Extracting Semantic Relations from Query Logs*"



Experimental evaluation

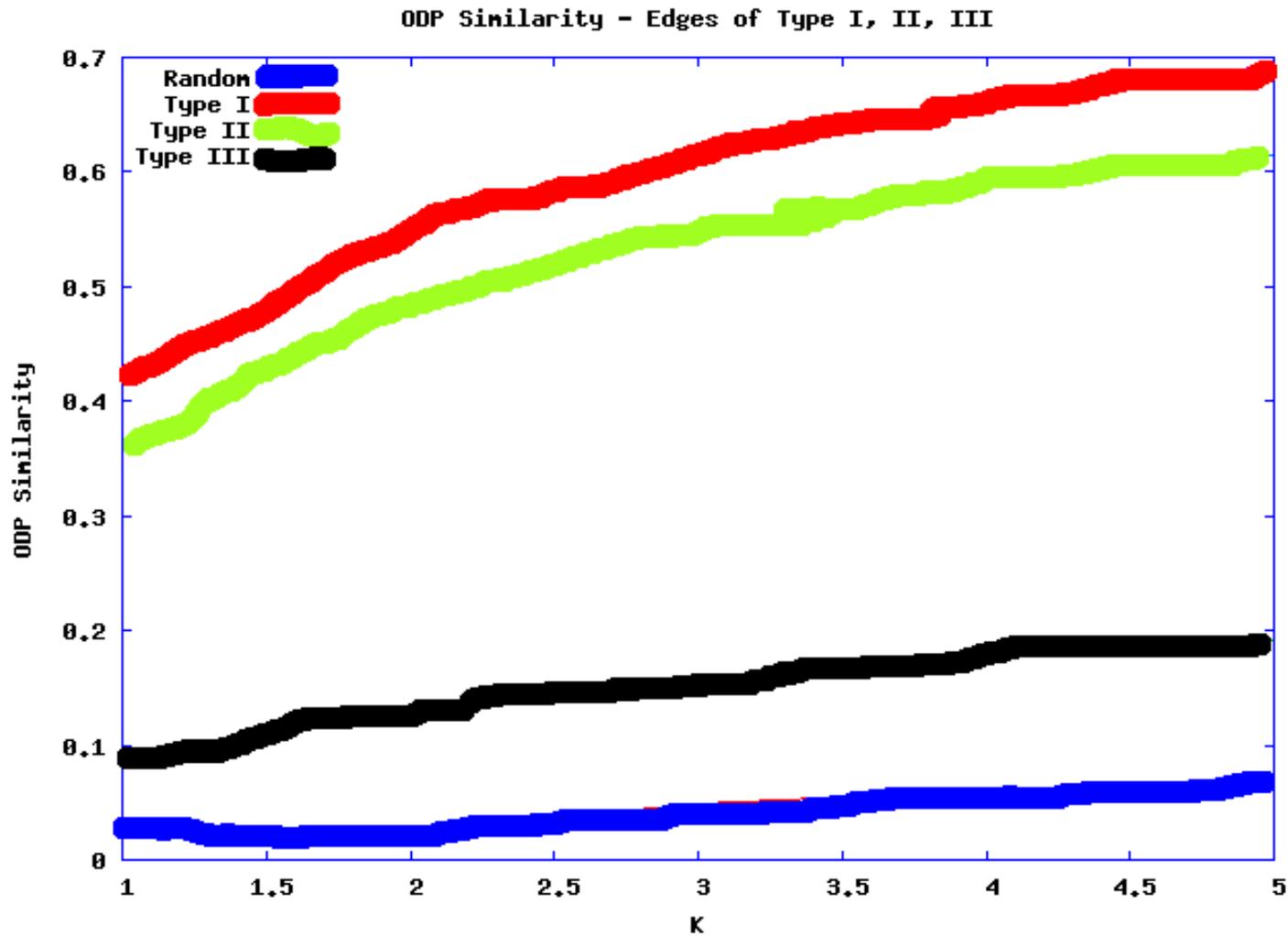
- We evaluated a 1000 thousand edges sample for each kind of relation
- We also evaluated a sample of random pairs of not adjacent queries (baseline)
- We studied the similarity as a function of k (the number of categories used)

[Baeza-Yates & Tiberi KDD' 2007] "Extracting Semantic Relations from Query Logs"



Experimental evaluation

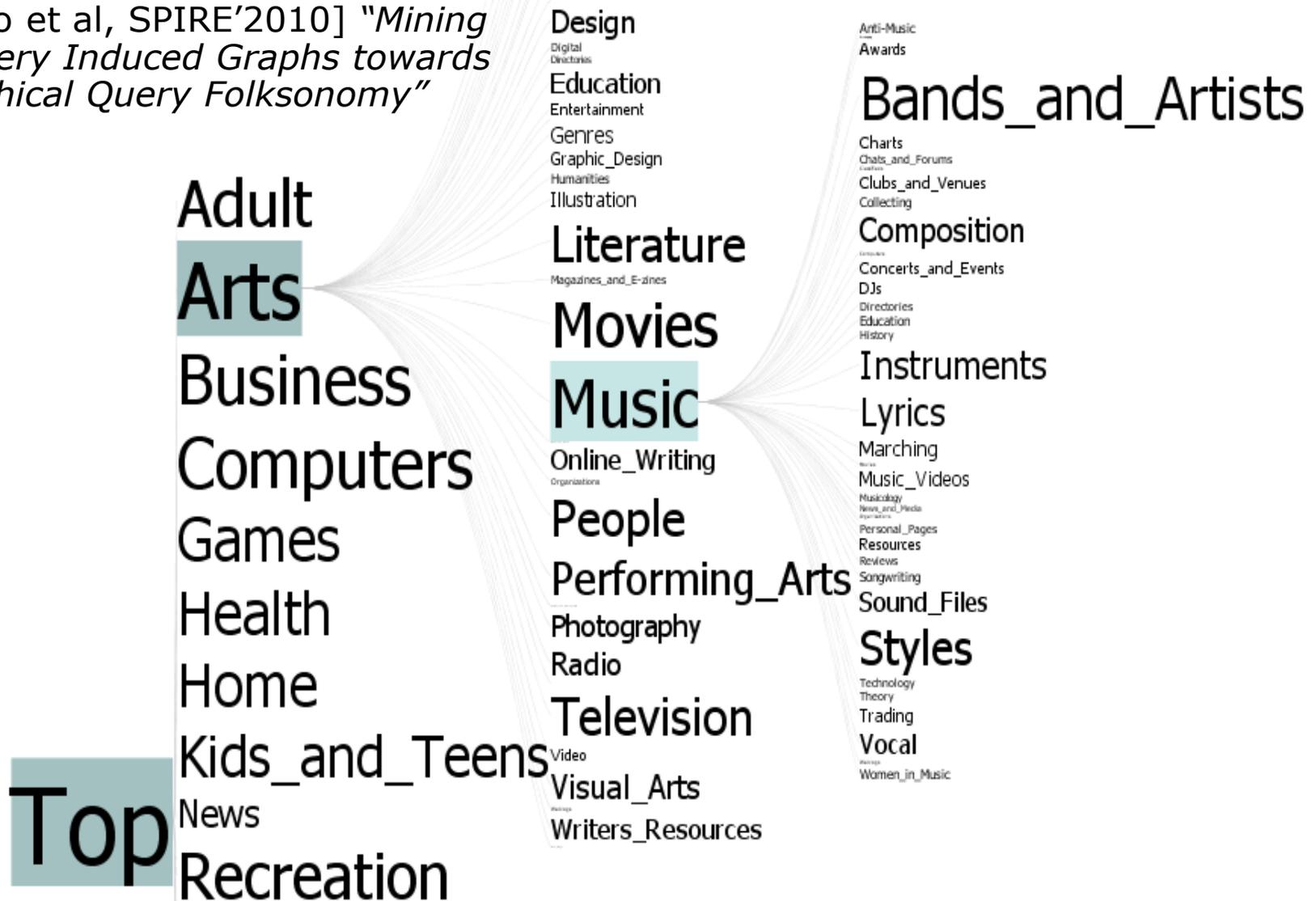
[Baeza-Yates & Tiberi KDD' 2007]
"Extracting Semantic Relations from Query Logs"

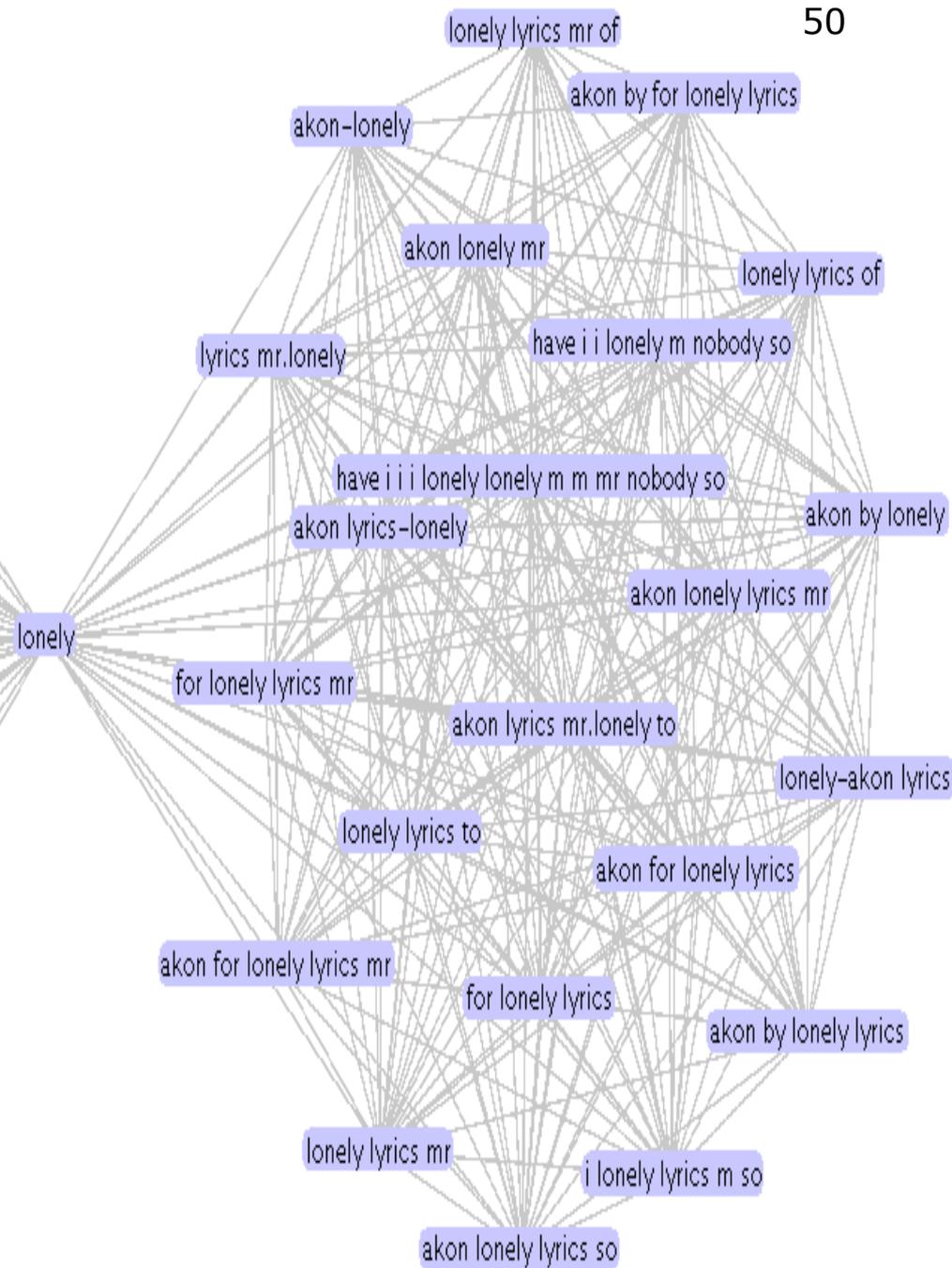




Mapping queries to ODP

[Francisco et al, SPIRE'2010] "Mining Large Query Induced Graphs towards a Hierarchical Query Folksonomy"







Hierarchical clustering

[Francisco et al, SPIRE'2010]
"Mining Large Query Induced
Graphs towards a Hierarchical
Query Folksonomy"

free

tommy_amour_jeff
r_us_toys
kementerian_malaysia_pendidikan
urban_hoax_legends
fax_internet_ofax
mvp_ducks_mcmurray
vida_guerra_phone
scuba_diving_snorkel
philippines_philippine_newspaper
ask_jeeves_www_ask
clit_bang_bangbus
philadelphia_wpvi_6
numa_dynemite_napoleon
gangs_gang_street
fry_frys_electronics
hobby_quarters_lobby
lolita_pedo_lolitas
ncaa_basketball_championships
paris_france_hotels
estate_real_banker
survivor_palaui_cbs
psalm_zine_poker
orleans_louisiana_new
friends_happy_tree
trisha_nude_actress
puppies_sale_for
plant_plants_light
auto_parts_autozone
cat_cats_punk
books_half_powell
book_camps_camp
nones none roman

free

pest_roach_bear
trails_forest_hiking
dota_warcraft_allstar
drinks_drink_bartending
aa_route_planner
□□□□_□□□_□□□
cena_john_wwe
lotto_lottery_super
massachusetts_boston_ma
outlet_springs_golf
bcb_spokane_bumiputra
poems_scholastic_children

candy_jelly_hershey
surveys_paid_jobs
fights_street_video
furniture_ikea_dining
jc_penney_penny
wine_baskets_sattui
preteen_models_lolita
sheet_music_piano
free cocks_black
fonts_font_free
tax_peas_state
model_modeling_models
cancer_breast_komen
ebay_motors_keyless
malaysia_of_kementerian
bed_billet_truck
constitution_amendment_amendments
pay_military_mypay
final_fantasy_advent
furniture_blinds_rugs
lyrics_mymp_moves
potter_harry_goblet
briggs_parts_stratton
navy_old_baby
lagu_pan_peter
gps_garmin_magellan
clothing_bealls_alloy
dade_miami_clerk

free

diet_recipes_exercise
battery_adapter_camcorder
guitar_tab_chords
amityville_horror_story
draft_nfl_mock
hair_prom_styles
encarta_terms_computer
county_harris_cook
smp_choppers_ride
typing_free_test
south_park_southpark
snoop_dogg_girls
books_michael_book
penguin_yeti_cubs
naruto_torrent_manga
wedding_patterns_shoes
creative_sound_blaster
court_supreme_cases
pool_tanning_swimming
test_permit_driving
vitamin_eas_liver
bluegreen_williamsburg_marriott
war_world_2
warcraft_world_wow
atlanta_baymont_braves
jazeera_al_aljazeera

free_games_yahoo

mortgage_homes_home
blue_book_value
inmate_county_orange
credit_report_free
card_credit_citi
sony_panasonic_motorola
states_directions_map
games_nick_nickelodeon
hair_hsn_qvc
karaoke_music_songs
dictionary_english_spanish
nike_vuitton_louis
sony_camera_canon
college_university_community
code_area_codes
hotels_marcus_neiman
inn_holiday_marriott
illinois_state_of
spyware_adware_lavasoft
free_chart_software
verizon_phone_telephone
victoria_secret_victorias
records_visa_state
exchange_currency_rate
download_music_p2p
homes_sale_houses
estate_real_insurance
airlines_york_west
pet_dog_foster
california_tax_state
movie_wars_star
chicago_sun_news
weather_newspaper_news
school_schools_high
screensavers_wallpaper_free
metro_bus_mta
lane_oreilly_o
lyrics_song_lyric
apartments_apartment_rent
music_mp3_download
music_top_billboard
iraq_navy_casualties
apprentice_news_nbc
free_sex_porn

Top





Some open issues

- Data Volume versus Better Algorithms
- Explicit versus implicit social networks
 - Any fundamental similarities?
- How to evaluate with (small) partial knowledge?
 - Data volume amplifies the problem
- User aggregation versus personalization
 - Optimize common tasks
 - Move away from privacy issues



Ultimate evaluation – on real traffic

- Examine users
- Conduct experiments
- Launch (by increments)



Predicting/analyzing clickthrough

- Key question: whether a user examined a specific position [Srikant et al. KDD'2010] "*User Browsing Models: Relevance versus Examination*"
- 3 models
 - Assume examination is independent of the other results for query
 - The **examination hypothesis**, [Richardson et al WWW'2007] "*Predicting clicks: Estimating the click-through rate for new ads*":
To be clicked, a result must be both examined and relevant
 - Assume examination depends on the pattern of clicks on prior results
 - Assume examination depends on both the pattern of clicks on prior results, and the relevance of prior results.



Examination depends on other results

- The **cascade hypothesis and model**, [Craswell et al. WSDM'2008] "*An experimental comparison of click position-bias models*"
 - Hypothesis assumes that users scan each result sequentially without any skips
 - Model further constrains that the user continues examining results until she clicks on a result, and does not examine any additional results after the click
- The **dependent click model**, [Guo et al, WSDM'2009] "*Efficient multiple-click models in web search*"
 - generalizes the cascade model to instances with multiple clicks
- The **user browsing model**, [Dupret et al, SIGIR 2008] "*A user browsing model to predict search engine click data from past observations.*"
 - In contrast, allows users to stop browsing the current results and instead reformulate the query



Examination depends on prior clicks and prior relevance

- The **click-chain model** [Guo et al, WWW'2009] "*Click chain model in web Search*"
 - If a user clicks on the previous result, the probability that they go on to examine more results ranges between 2 values depending on the relevance of the previous result
- The **general click model** [Zhu et al, WSDM'2010] "*A novel click model and its applications to online advertising*"
 - Treats all relevance and examination effects in the model as random variables
- All the above models relate to "perceived" relevance – whether the user considers the result relevant before he clicks on the result
- There also exist "post-clicks" models

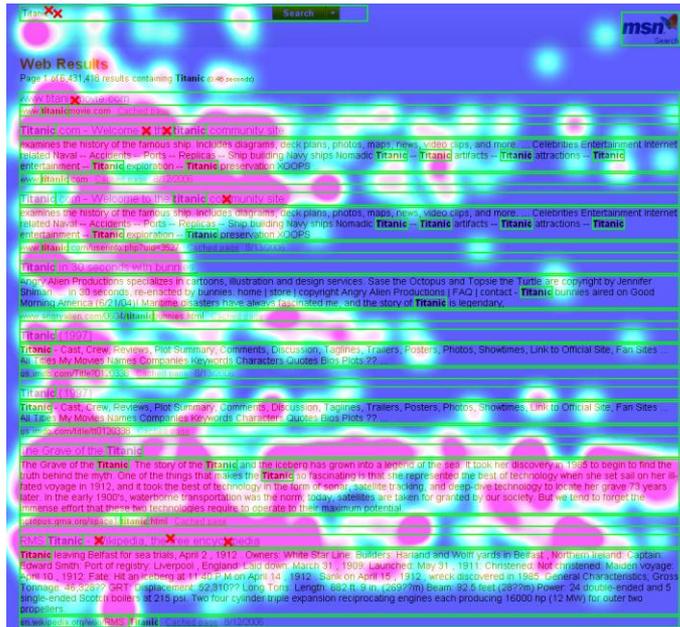


Post-clicks models

- The **dynamic bayesian model** [Chapelle et al. WWW'2009] "*A dynamic bayesian network click model for web search ranking*"
 - Uses the "user satisfaction" (post-click relevance) of the preceding click to predict whether the user will continue examining additional results
- The **session utility model** [Dupret et al., WSDM'2010] "*A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine*"
 - Proposes a user browsing model based on the "intrinsic" (post-click) relevance of the sequence of clicked results in a user session
- Srikant et al. examined all these models and proved that: "*relevance of the result for that query instance is strongly correlated with clicks on other results and is responsible for a substantial portion of the changes in conditioned CTR*"

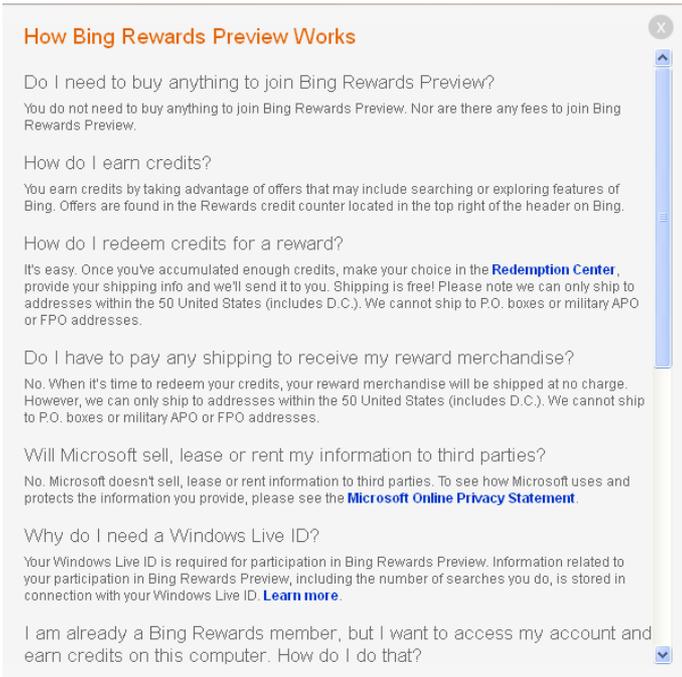


Experimentation models



Enroll Beta-testers

- Limited in scope
- Has a cost



Observe users in users-studies

- Eyetracking, Mouse tracking
- [Granka et al., SIGIR'2004] "Eye-Tracking Analysis of User Behavior in WWW Search"
- [Cutrell et al., CHI'2007] "What are you looking for?: an eye-tracking study of information usage in web search"
- Etc.





Experiment and launch

- Bucket AKA 1% experiment AKA A/B testing
 - [Kohavi et al, Journal of Data Mining and Knowledge Discovery, 2009] *“Controlled experiments on the web: survey and practical guide”*
- A key advantage of large scale – our testers are real users
 - Set aside a random (or not so random) % of users identified by cookies, i.e. a fraction of real traffic, and deploy the chosen feature just for them for a week or two
 - Measure multiple metrics during this experiment
 - The key is in interpreting the metrics and making the smart choice, a small decrease in revenue might be acceptable if user’s happiness is increased for instance
- For major features, launch in increments to verify that performance is maintained
- Remember to keep “hold back” experiments to keep improving



Explicit user aspects

Session 3

TM

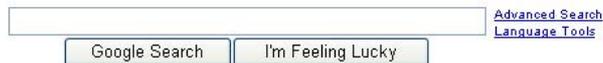
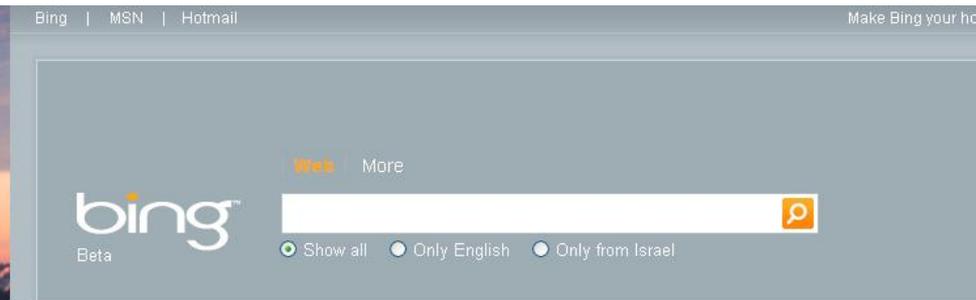
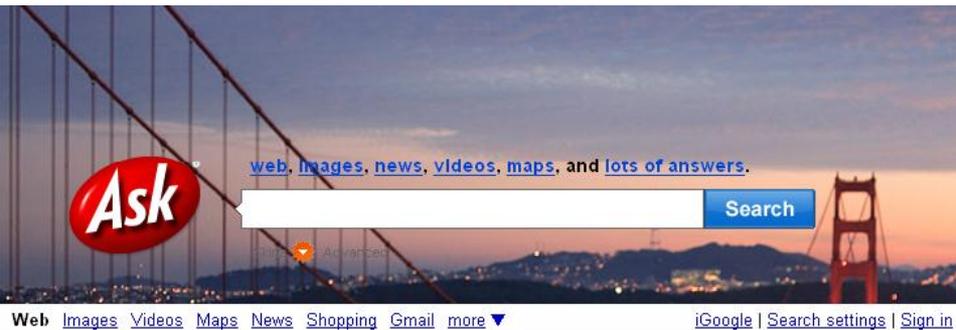


Search rectangle

Very little differences between major search engines
A rectangle – text box for your queries

Other forms of rectangles?

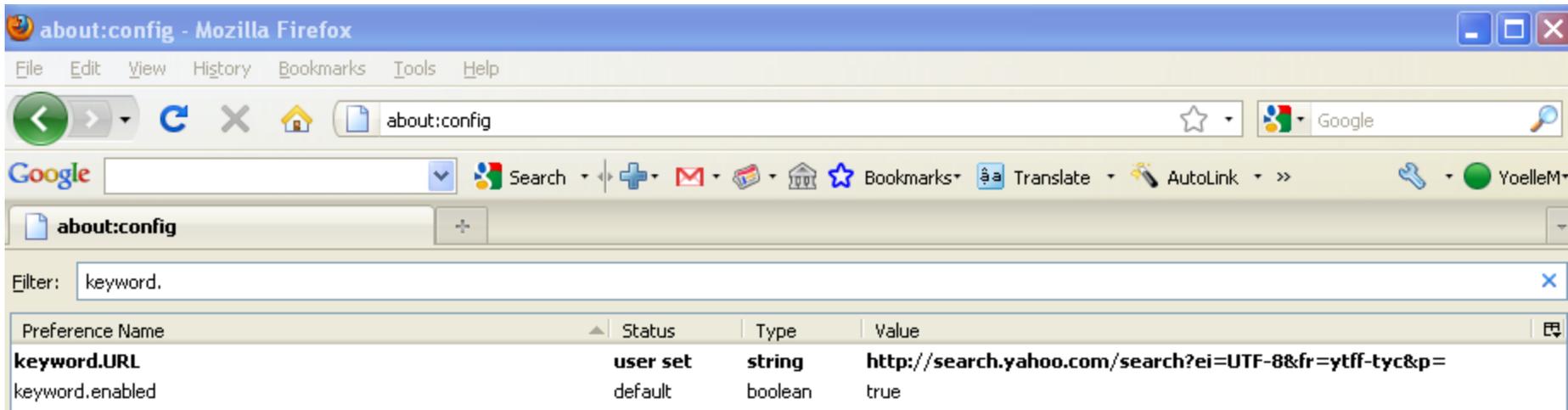
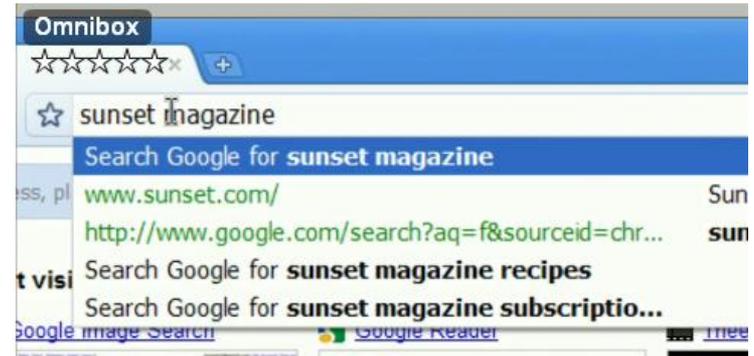
- Embedded in a portal
- Always here in a toolbar
- Ultimate rectangle: omnibox





Omniboxes

- Ultimate version in Google chrome
 - Merges the functionality of the address and the search boxes.
 - Browser decides whether the user's intent is navigation or search
- Firefox since release 2.0
 - When entering a word entered in the address bar triggers Google's "I feel lucky"
 - Customizable feature: By modifying the `keyword.URL` property (accessed by typing `about:config` in the address bar





Web query languages

- A few facts
 - There is no standard Web query language – semantics are not agreed upon, engines reserve the right to change their interpretation
 - Stemming?
 - Conjunction or not?
 - Query rewrite?
- Free-text query format won!
 - Queries are whatever users enter
 - Controlled vocabulary approach is dead
 - Spelling mistakes are acceptable
 - Difference with natural language?
 - Difference between a query and a question?



Common query operators

Operator Syntax	Details	Google	Yahoo! Search	Bing	Ask
“.” double quotes surrounding a string	Phrase search	yes	yes	yes	yes
+ preceded by a space, operates on the term/phrase that immediately follows	This operator ensures that the associated term is included “as is” in the results	yes	yes	yes	yes
- preceded by a space, operates on the term/phrase that immediately follows, Bing uses NOT as well	This operator ensures that the associated terms do not appear in any result	yes	yes	yes	yes
OR (as well as) operates on preceding and succeeding terms or phrases	Equivalent to a Boolean OR	yes	yes	yes	yes
site: Followed by a site name	Returns results from the specific site only	yes	yes	yes	yes
hostname: Followed by a host name	Returns results from the specific host only	no	yes	no	yes
url: Followed by a URL	Checks that the following url exists in the engine index	no	yes	yes	no
inurl: Followed by a term	Returns results whose URL contains the specified term	no	yes	no	yes
intitle: Followed by a term	Returns results whose title contains the specific term	no	yes	yes	yes
inlink:/inanchor: Followed by a term	Returns results that contain the specific term in their link or anchor metadata	yes	no	yes	yes

Table 2.3: Common query operators

[Baeza-Yates & Maarek, Modern IR 2nd edition, 2011] *“Web Retrieval” chapter*



Not so common operators

- **Google**
 - * should be treated “as a placeholder for any unknown term”
- **Yahoo**
 - **link: <url>** returns documents that link to a specific url, a feature provided by the Yahoo! Site Explorer tool
 - Other shortcuts retrieved by **!list** in the search rectangle include **!news**, **!flickr**, **!wiki**, **!map**, etc.
- **Bing**
 - **AND/&** interestingly enough, Bing provides this operator even if it claims that “by default all searches are AND searches”.
 - **()** used to group words in conjunction with other operators (-/+).
 - **filetype:**, **contains:**, **ip:**, **feed:**, **prefer:** etc.
- **Ask**
 - **afterdate:**, **beforedate:** followed by a date (yyyymmdd)
 - **betweendate:** followed by 2 dates separated by a .
 - **last:** followed by a given time period, among {**week**, **2weeks**, **month**, **6months**, **year**, **2years**}



Dynamic query suggestion

- Query suggestions appear as you type in rectangle
- Some history
 - Google labs in 2004
 - Google toolbar in 2006
 - yahoo.com, search.yahoo.com in 2007
 - youtube.com and google.com in 2008



inform

[Advanced Search](#)
[Language Tools](#)

informative **speech topics**
information **technology**
informatica
informal **wedding dresses**
information**week**
informed **consent**
informatics
information **assurance**
information **systems**
information

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

inform|

information
informative speech topics
informatics
webmd health information
information technology

[Options](#) ▾



© 2009 Yahoo! [Privacy](#) / [Legal](#) - [Submit Your Site](#)





Y! Search assist: Don't stop at query completion

YAHOO! [Web](#) [Images](#) [Video](#) [Local](#) [Shopping](#) [News](#) [More](#) [Search](#) [Options](#)

- hong kong disneyland
- hong kong hotels
- yahoo hong kong
- hong kong airport
- hong kong map

Explore related concepts:

- singapore
- taiwan
- macau
- china
- japan
- shanghai
- tokyo
- beijing

[Settings](#)

YAHOO! [Web](#) [Images](#) [Video](#) [Local](#) [Shopping](#) [News](#) [More](#) [Search](#) [Options](#)

- hong kong disneyland
- hong kong map
- hong kong actress
- hong kong flag
- hong kong phoey

-  ... disneyland
-  hong kong map
-  ... actress
-  hong kong flag



Even ads

hilton hotels [Advanced Search](#)

[Hotels by Hilton - Hotel Reservations, Deals, and Room Rates](#) [Preferences](#)

www1.hilton.com/ [Language Tools](#)

hilton hotels **corporation**

hilton hotels **careers**

hilton hotels **hawaii**

hilton hotels **new york**

hilton hotels **chicago**

hilton hotels **washington dc**

hilton hotels **jobs**

hilton hotels **florida**

hilton hotels **corp**

hilton hotels **london**

[Hilton.com Official Site](#) Sponsored Link

www.Hilton.com Book direct on **Hilton.com** For Our Best Rates, Guaranteed.



... and universal search

weather haif

weather haifa
weather haifa **10 day**
weather haifa **november**
weather **halifax**
weather haifa **october**

Weather: 54°F in Haifa, Israel

 Sun 55° 71°	 Mon 53° 71°	 Tue 50° 64°	 Wed 51° 66°
--	--	---	--

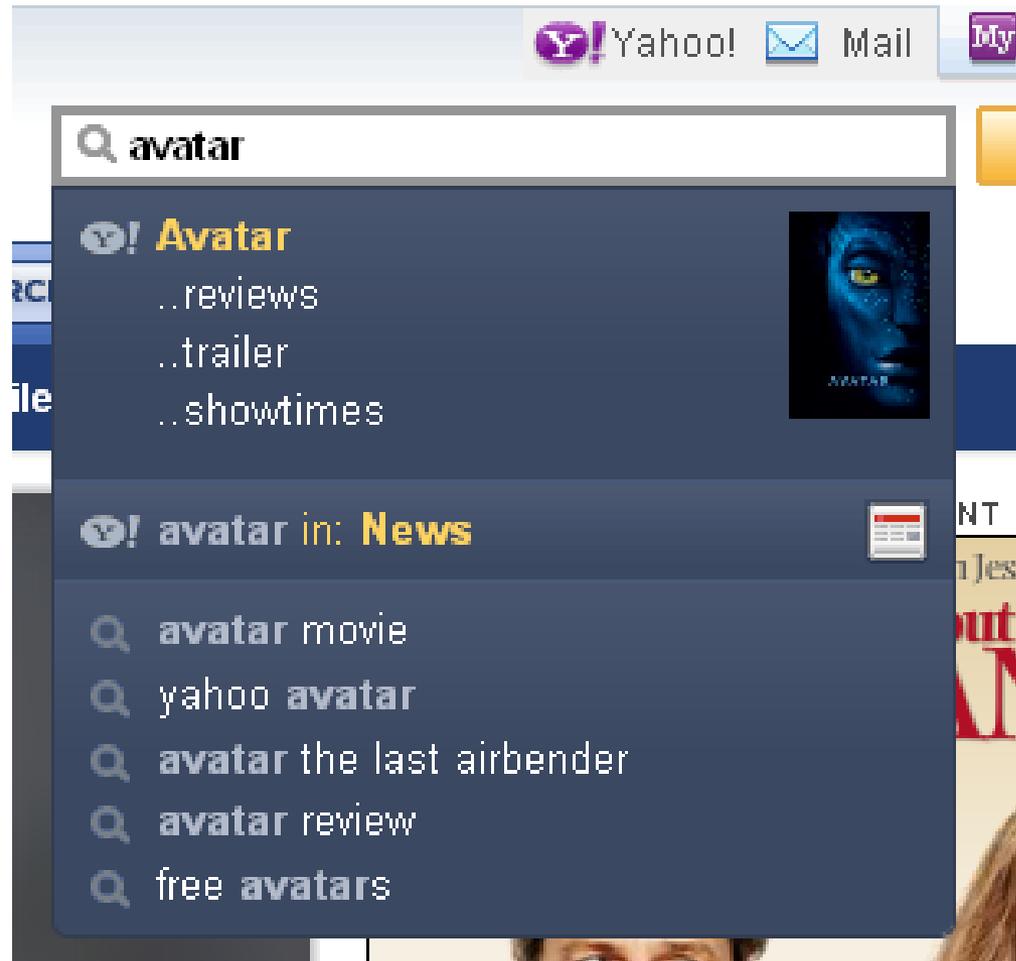
The Official Google Blog.

Two new features enhance search beyond the results page, 12/11/2009 10:16:00 AM

<http://googleblog.blogspot.com/2009/12/two-new-features-enhance-search-beyond.html>



..and always more





Dynamic query suggestions

- Key difference between dynamic query suggestions and query assistance?
 - Query suggestions take as input a prefix as opposed to a full query
- Challenges
 - Limited input – prefix only
 - Latency
 - Freshness
 - Locality
 - Diversity (danger of rich gets richer syndrome)
 - etc.



Parallel with regular search

more critical

Effectiveness

	Query x Docs	Prefix x Queries
Efficiency	~1s	Typing pace
Automatic spelling correction	Well spelled/formed queries	Well spelled/formed queries
Filtering	Inappropriate documents	Inappropriate queries
Freshness	Fresh documents (news)	Fresh queries
Personalization	Personalized ranking	Favorite queries
Local	Promote local results	Promote local queries pizzeria in ...
Universal	Merge results from different sources	Merge queries from different sources (fresh vs local?)



The "Voice of the search engine"???

tony blair is

tony blair is **the antichrist**
 tony blair is **catholic**
 tony blair is **dead**
 tony blair is **a catholic**

berlusconi is |

berlusconi is **not speaking in my name**

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

berlusconi is |

berlusconi is magic

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

sarkozy is |

sarkozy is jewish
 sarkozy is a jew
 sarkozy is jew
 sarkozy is short
 sarkozy is greek

obama is |

obama is **the antichrist**
 obama is **an idiot**
 obama is **a socialist**
 obama is **a liar**
 obama is **irish**
 obama is **a muslim**
 obama is **not black**
 obama is **a joke**
 obama is **a communist**
 obama is **nailin paylin**

Sponsored Links

[Is Obama The Antichrist?](#)
 The Antichrist Revealed - When?
 Biblical Perspective On Last Days
[InvitationToChrist.org](#)

bill clinton is

bill clinton is is
 bill clinton is **quote**
 bill clinton is **black**
 bill clinton is **sexy**
 bill clinton is **the first black president**
 bill clinton is **a pimp**
 bill clinton is **a rockefeller**
 bill clinton is **impeached**
 bill clinton is **the antichrist**

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

hillary clinton is |

hillary clinton is evil
 hillary clinton is the devil
 hillary clinton funny
 hillary clinton is a communist
 hillary clinton is a witch



Not only with people names ...

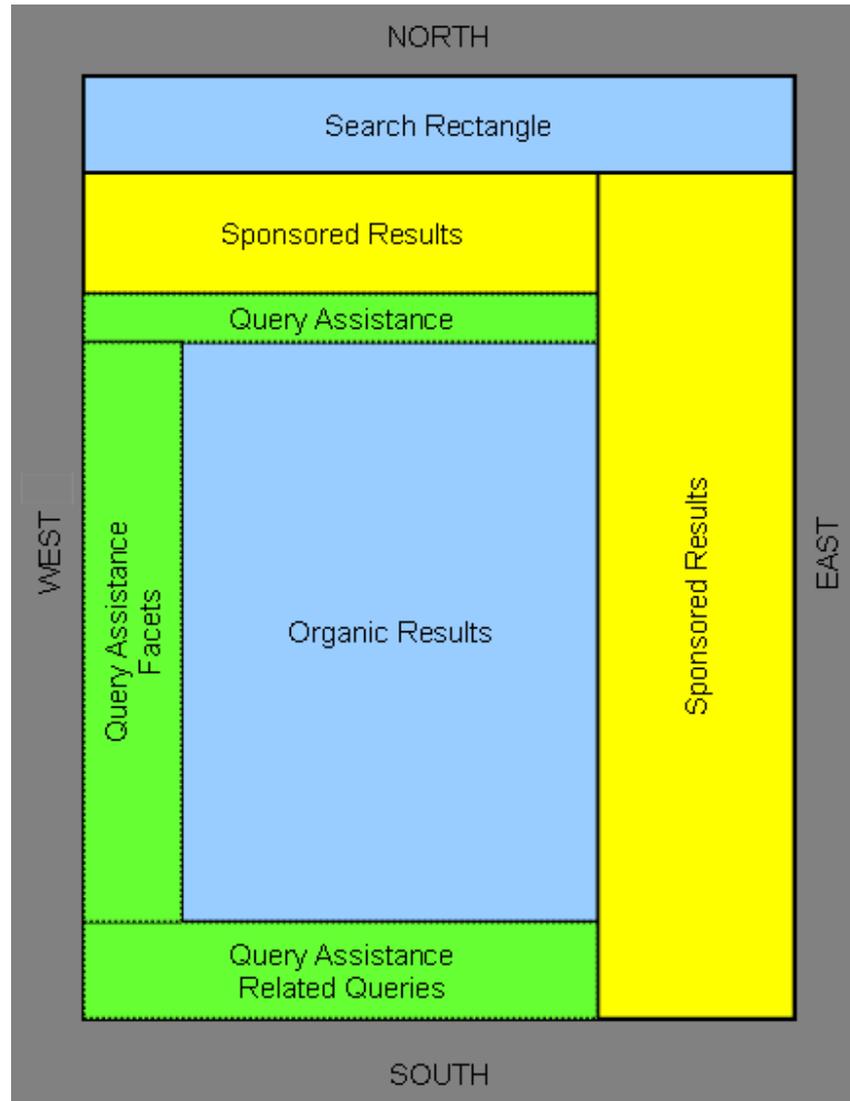
[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

sony is	Search
sony is evil	
sony is 10	
sony is storage	
sony is giving away psp consoles free	
sony is the best	

nokia is
nokia is from what country
nokia is from
nokia is the best
nokia is from which country
nokia is based in
nokia is of which country
nokia is jewish
nokia is owned by
<input type="button" value="Google Search"/> <input type="button" value="I'm Feeling Lucky"/>



SERP – basic layout





Google SERP

Search rectangle

Organic results

Sponsored results

Query Assistance

Web Images Videos Maps News Shopping Gmail more - Yoelle Maarek -



X
Search

About 62,100,000 results (0.24 seconds) Advanced search

Everything

Images

Videos

News

Shopping

Realtime

More

Any time

Latest

Past 24 hours

Past 2 days

Past week

Past month

Past year

▶ [Hotels in Beijing - Book your Beijing hotel. | hotels.com](#) Ad

Great prices available.
www.hotels.com

Beijing China maps.google.com










Hotels - Restaurants - Summer Palace - Temple of Heaven - Forbidden City - Tiananmen Square - Beijing Hutong - Wangfujing Street

Beijing - Wikipedia, the free encyclopedia

Location of **Beijing** Municipality within **China**. Coordinates: 39°54'50"N 116°23'30"E

Featured Jobs: China **Beijing, China**. Content Analyst, English Speaking, Fundamentals Beijing Team. Contractor - M Content Invest Research Cont ... [careers.thomsonreuters.com](#) > Careers - [Cached](#)

 Nicholas Creswell shared this

Searches related to beijing, china

[great wall of china](#)

[china map](#)

Ads

150 Hotels in Beijing

Book your Hotel in **Beijing** online. Save up to 50% on your reservation! booking.com is rated ★★★★★ [www.booking.com/Beijing](#)

Beijing China

Visiting **Beijing**? Find Deals & Read Hotel Reviews! tripadvisor.com is rated ★★★★★ [www.tripadvisor.com/Beijing](#)

See your ad here >



Search rectangle

Organic results

Sponsored results

Query Assistance



Web Images Video Local Shopping News Apps More

Read our updated data policy

beijing, china

Search

Options

QuickApps

SafeSearch - On

29,100,000 results for beijing, china

Related Points of Interest



Forbidden City



Great Wall



Temple of Heave...



Tiananmen Squar...



Chairman Mao Me...



Dashanzi



Panjiayuan



Xidan



Zhengyangmen



Shichahai

Beijing Hotels

Sponsored Results

Plenty of choices at **Beijing** Hotels. Save up to 75%. Don't miss!
www.agoda.com/Beijing-Hotels

Beijing Travel Advice

Find Hidden Gems, Get Visa@ Money Tips And Locate ATM's In **Beijing!**
www.tripadvisor.com/visa

250 Hotels in Beijing

Book your Hotel in **Beijing** online. No reservation costs. Great rates!
Booking.com/Beijing-Hotels

Discount China Tours

100+ **China** tour packages starting fr \$450. Book & get discount now!
www.china-tour.cn

Beijing - Visitor Guide

travel.yahoo.com

Overview

Hotels

Restaurants

Current Local Time

7:29 pm (CST)

Weather Forecast

Today Partly Cloudy 93°F | 72°F

Tomorrow PM Thunderstorms 90°F | 72°F

Thursday Thunderstorms 89°F | 67°F

[Extended forecast on Yahoo! Weather](#)



Beijing - Wikipedia, the free encyclopedia

[Etymology and names](#) | [History](#) | [Geography](#) | [Cityscape](#)

Beijing, also known as Peking, is a metropolis in Northern **China**, and the capital of the People's Republic of **China**. Governed as a municipality under direct administration of the...

en.wikipedia.org/wiki/Beijing - [Cached](#)

[Beijing Official Website International - eBeijing.org](#)

Luxury China Tours

Luxury tours to **China**, Beijing, Xian, Shanghai, Guilin, Yangtze River.
www.HQChinaTours.com

China Silk Tours

Over 10 years, America's Leading **China** & Southeast Asia Tour Operator
www.chinasilktours.com

Top Beijing Travel Agency

Best **Beijing** Tours, **Beijing** Travel Guide, **Beijing** China Travel
www.topbeijingtravel.com

guide and information

high quality service and guide of major attractions—CITS
www.chinabravo.com

Beijing, China Hotels

Low Rate Guarantee. Free In-Room Internet & Hot Breakfast. Book Now!
www.hiexpress.com

2011 Special China Tours

Selected Destinations, frn \$414; Incl. **Beijing**, Xian, Tibet, Yangtze.
ChinaTourGuide.com/China_Tour

More Sponsors:

[beijing, china](#)

[See your message here...](#)



Bing SERP

Search rectangle

Organic results

Sponsored results

Query Assistance

Web Images Videos Shopping News Maps More MSN Hotmail Sign in Rewards Haifa, Hefa Preferences

bing

Web Places Images Blogs Videos More

Search: beijing, china

1-10 of 28,800,000 results · Advanced

ALL RESULTS

Beijing Travel Advice Ads
www.tripadvisor.com/visa · Find Hidden Gems, Get Visa® Money Tips And Locate ATM's In **Beijing!**

250 Hotels in Beijing
Booking.com/Beijing-Hotels · Book your Hotel in **Beijing** online. No reservation costs. Great rates!

Discount China Tours
<http://www.china-tour.cn> · 100+ **China** tour packages starting fr \$450. Book & get discount now!

China Silk Tours
www.chinasilktours.com · Over 10 years, America's Leading **China** & Southeast Asia Tour Operator

Other ideas: [beijing china travel](#) · [beijing china tours](#)

Beijing, China Travel Guide BING TRAVEL

Explore top attractions and photos. Find great deals on flights and hotels.



91°F Sunny (Clear)

Attractions: [Great Wall](#) · [Yonghe Gong \(Lama Temple\)](#) · [Forbidden City](#) · [See all](#)

Private China Tours
 Private tours to **China**, Beijing, Xian, Shanghai, Guilin, Yangtze River.
www.HQChinaTours.com

Top Beijing Travel Agency
 Best **Beijing** Tours, **Beijing** Travel Guide, **Beijing China** Travel
www.topbeijingtravel.com

guide and information
 high quality service and guide of major attractions—CITS
www.chinabravo.com

Holiday Inn Express Hotel
 Free Internet and Breakfast! in **Beijing, China**. Official Site.
www.hiexpress.com

2011 Special China Tours
 Selected Destinations, frn \$414; Incl. **Beijing**, Xian, Tibet, Yangtze.
ChinaTourGuide.com/China_Tour

Weather Beijing China **Temperature Beijing China**
Facts about Beijing China **Great Wall of China Beijing**
Beijing China Map **Beijing China Travel**
Beijing Time **Beijing City Map**

Beijing Travel Advice Ads
www.tripadvisor.com/visa · Find Hidden Gems, Get Visa® Money Tips And Locate ATM's In **Beijing!**

250 Hotels in Beijing
Booking.com/Beijing-Hotels · Book your Hotel in **Beijing** online. No reservation costs. Great rates!



Traditional organic results

- Regular result
 - [Title](#)
 - Snippet computed on the fly, query dependent, highlights the **right keywords**
 - [URL](#)
- Where do we get titles from?
- How to build snippets?
 - Identify shortest window
 - Use information from parents' page for inferring titles/snippets?

[Amitay & Paris, CIKM'2000]
*"Automatically Summarising Web Sites,
Is There A Way Around It?"*

[Beijing Official Website International - eBeijing.gov.cn](#)

eBeijing: **Beijing's** Homepage, the Official English website of the City of **Beijing** ... I want to learn about **China's** education system. I want to learn about education ...
[www.ebeijing.gov.cn](#) - [Cached](#)

[Beijing Official Website International - eBeijing.gov.cn](#) 🔍

eBeijing: **Beijing's** Homepage, the Official English website of the City of **Beijing**.
[Beijing Info - Essence of Beijing](#) - [Beijing Travel Planner](#) - [Guide to Beijing Hospitals](#)
[www.ebeijing.gov.cn/](#) - [Cached](#) - [Similar](#)

[Beijing Official Website International - eBeijing.gov.cn](#)

eBeijing: **Beijing's** Homepage, the Official English website of the City of **Beijing** ... I want to learn about **China's** education system. I want to learn about education ...
[www.ebeijing.gov.cn](#)



More about results

- Results from a same site
 - Earlier in the game: indents
 - Now navigational shortcuts aka Sitelinks/Quick links
 - “navigational shortcuts [...] are displayed below the website homepage on a search results page and let users directly jump to selected points inside the website”
 - Not trivial as the goal is to maximize the benefits for a majority of users, while showing only relevant links in a limited real estate.
- Will keep improving since the best links are inferred from learning users' behavior via clicks and toolbar data

▶ [Delta Air Lines - Airline Tickets and Airfare to Worldwide ...](#)  

[www.delta.com/](#) - Cached

Airline tickets, flight schedules, and air travel deals to destinations worldwide at **Delta.com**.

 [Show stock quote for DAL](#)

Book a flight	Flight Schedules
Flight Status & Updates	Customer Care
SkyMiles	Travel Information v
Baggage	Manage My Account

[More results from delta.com »](#)

[Chakrabati, Kumar & Pundera, WWW'2009] "Quicklink selection for navigational query results"



Oneboxes/Direct Display

- **What are these?**

- Very specific results (rather than links), answering very precise queries with a unique answer
- Displayed above regular Web results, due to their high relevance, and in a slightly different format.

- **How does it work**

- Hack! Triggered by specific terms in the user's query that indicate a clear intent
- Ultimate search result experience but only in very specific cases.

- **Examples:**

- Weather, Movie schedules, ??

bing Web

wimbledon scores

Web Blogs More

ALL RESULTS

[Wimbledon](#) Wimbledon, Jun 20 - Jul 3, 2011

Men's Tennis > Women's Tennis

ROUND	DATE	PLAYER1	PLAYER2	RESULT
Rnd 4	Jun 27	R Nadal (w)	J Potro	7-6(6) 3-6 7-6(4) 6-4
Qtrs	Jun 29	N Djokovic (w)	B Tomic	6-2 3-6 6-3 7-5
Qtrs	Jun 29	J Tsonga (w)	R Federer	3-6 6-7(3) 6-4 6-4 6-4
Qtrs	Jun 29	R Nadal (w)	M Fish	6-3 6-3 5-7 6-4
Qtrs	Jun 29	A Murray (w)	F Lopez	6-3 6-4 6-4

Rankings · FOX Sports on MSN (STATS, Inc. © 2011)

YAHOO!

Web Images Video Local Shopping News Apps More

weather beijing

[Extended weather forecast for Beijing, China](#) (Change Location)

weather.yahoo.com

<p>Currently: 88°F Mostly Cloudy 90°F 75°F</p>	<p>Tomorrow PM Thunderstorms 90°F 74°F</p>	<p>Saturday Thunderstorms 87°F 71°F</p>
--	--	---

[Try this search on my phone](#)

bing Web

Flights from **Tel Aviv-Yafo, Israel** to **Beijing, China**

FROM TO

LEAVE RETURN [Find flights](#)

bing.com/travel

Google

flights from tel aviv to beijing

About 8,060,000 results (0.20 seconds)

Flights from Tel Aviv, Israel (TLV) to Beijing, China (all airports)

Non-stop flights: 3 per week, 9h 35m duration

Airlines: [EI AI](#)

[+ Schedule of non-stop flights](#)



Structured display by agreement

Publishers and search engines agree on a preferred display format

- Yahoo!: Search Monkey in 2007
 - [Mika, SemSearch'2008] Microsearch: An interface for semantic search"
- Google: Rich snippets launched in 2009
 - <http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=99170>
- Also on Bing

The image shows a screenshot of a Yahoo! search for 'china'. The search bar contains 'china'. Below the search bar, there are sections for 'QuickApps', 'SafeSearch - On', and search results for 'china' (354,000,000 results). A 'Also try:' section lists related terms like 'china flood 2011', 'china airlines', and 'china ufo'. The main result is 'China - Wikipedia, the free encyclopedia', which has structured snippets for 'Etymology', 'History', 'Territory and environment', and 'Economy'. Below this is the 'Search Monkey' logo, a cartoon monkey wearing glasses. Below the logo, there are four vertical lines that connect the structured snippets from the Wikipedia result to the corresponding snippets in the Bing search results shown below. The Bing search results for 'china' also show the same structured snippets for the Wikipedia result, demonstrating a consistent display format across different search engines.



Interactive layouts

- Bing "More on this page"
- Google Instant Preview
- Yahoo browse in place

[Beijing - Wikipedia, the free encyclopedia](#)
 Etymology and names · History · Geography · Cityscape
Beijing, also known as Peking, is a metropolis in Northern China, and the capital of the People's Republic of China. Governed as a municipality under direct ...
en.wikipedia.org/wiki/Beijing

More on this page

PAGE SECTIONS

1. [Etymology and names](#)
"Beijing" means "Northern Capital", in line with the common East Asian tradition whereby ...
2. [History](#)
The earliest remnants of human habitation in the Beijing municipality

[China - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/China - Cached

China is seen variously as an ancient civilization extending over a large area in a nation and/or a multinational entity ...

People's Republic of China	Religion in China
History of China	Province (China)
Economy of the People's Republic of ...	Chinese language
Geography of China	Outline of China

More results from wikipedia.org >

For the state commonly known as **China**, see **People's Republic of China** .

With nearly 4000 years of continuous history, **China** is one of the world's oldest ...

Similar

[Beijing - Visitor Guide](#)
travel.yahoo.com

Overview Hotels Restaurants

Current Local Time
7:51 pm (CST)

Weather Forecast

Today	Mostly Cloudy	90°F 75°F
Tomorrow	PM Thunderstorms	90°F 74°F
Saturday	Thunderstorms	87°F 71°F

[Extended forecast on Yahoo! Weather](#)





Once a query is issued, the users' needs (informational, navigational as well transactional) can be either:

- satisfied
 - Users get their answer immediately from a onebox result, such as calculator, weather, sports results, etc.)
 - Almost immediately after they click on one or a few of the top results
- partially satisfied
 - Users have undertaken a “research task”, no single Web page holds all the needed information.
 - Needs susceptible to trigger research tasks: travel needs, homework, education needs, or health information
- not satisfied at all
 - Users did not formulate their query well or
 - Relevant content simply does not exist.

We focus here on research tasks and reformulation tools



Help users reformulate their queries

Related Queries

1. Content-aware approaches: Use SERP or target pages to measure query similarity
2. Content-ignorant approaches: Use Clicks
3. Query-flow approaches: Monitor the users' sequential search behavior to better understand query intent

The screenshot shows a Yahoo! search results page for the query "villas in italy". The page includes the Yahoo! logo, search filters (Web, Images, Video, Local, Shopping, More), and a search bar containing the query. Below the search bar, there are options for "Search Pad" and "SearchScan - On". The search results show 22,400,000 results for "villas in italy". A highlighted result is "Destination Villas Italy Vacation Rental" from "www.destination-villas.com". Below this, there is a result for "Italian Villa Rentals and Italy Vacation Rentals - The Parker Company" from "www.parkervillas.com/". At the bottom of the search results, there are "Searches related to villas in italy" including "villas for sale in italy", "map of italy", "hotels in italy", "villas in spain", "italian villas", "owners direct", "first choice", and "thomas cook". The page also features a "Go" button with a progress indicator and the number "1" selected.



Dynamic results *as you type*

- Google Instant
 - Leverages Google Auto-completion
 - One dimensional
- Yahoo! DirectSearch
 - Leverages SearchAssist
 - Two dimensional

beijing

beijing weather
beijo
beijing time
beijing airport

About 337,000,000 results (0.19 seconds)

[Beijing China](#) maps.google.com



[Hotels](#) - [Restaurants](#) - [Summer Palace](#) - [Temple of Heaven](#) - [Forbidden City](#) - [Tiananmen Square](#) - [Beijing Hutong](#) - [Wangfujing Street](#)

[Beijing](#) - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/**Beijing** - Cached

Beijing also known as Peking is a metropolis in Northern China, and the capital of the People's Republic of China. Governed as a municipality under direct ...

[Etymology and names](#) - [History](#) - [Geography](#) - [Cityscape](#)

beij

beijing bullet
beijing olympic 2008
beijing olympics
beijing weather
beijing china
beijing 2008 olympic games
beijing airport
beijo bags
beijing hotels
beijing time

BEIJO BAGS

TOP SITES

[Beijo Bags: Handbags | Accessories | Home Based...](#)
 www.beijobags.com

[Buy Beijo Handbags on Sale: Buy and Review |...](#)
 chic-handbags.com/beijo-handbags

Collection
 www.beijobags.com/bc_0.aspx



Actionable results

- Cached
 - Translate
 - Similar
- } most
- Demoted features
 - Searchwiki>Stars in Search (Google)
 - Notepad (Google) Search Pad (Yahoo! Search)
 - Recent features
 - Google +1
 - Yahoo! Apps

PS: we discuss here only actionable “results” as opposed to actionable web pages through services offered by toolbars or browsers (e.g., sidewiki out of scope)



From Searchwiki ...

Web [Images](#) [News](#) [Translate](#) [Scholar](#) [YouTube](#) [Gmail](#) [more](#) ▼

Google

Search: the web pages from Israel

Web [+ Show options...](#)

[Haifa - Wikipedia, the free encyclopedia](#)

Haifa, built on the slopes of Mount Carmel, has a history dating back to **Haifa** is Israel's third-largest city, consisting of 103000 households. ...

[Etymology](#) - [Early history](#) - [Modern history](#) - [Demographics](#)
en.wikipedia.org/wiki/**Haifa** - [Cached](#) - [Similar](#) -

Type your comment here

[Haifa Wehbe - Wikipedia, the free encyclopedia](#)

Haifa Wehbe was born in Mahrouna, a small Shiite farming town in Southern ... **Haifa Wehbe** has also performed with artists from outside the Arab world. ...

en.wikipedia.org/wiki/**Haifa_Wehbe** - [Cached](#) - [Similar](#) -

Goooooooooooooogle
1 2 3 4 5 6 7 8 9 10 [Next](#)

- User's feedback
 - Comment
 - Promote
 - Remove





To "Stars in Search"

[Haifa Israel](#) maps.google.com



[Haifa](#) - [Wikipedia, the free encyclopedia](#) ☆

Haifa (Hebrew: חיפה About this sound Heifa^(help·info); Arabic: حَيْفَا About this sound Hayfā^(help·info)) is the largest city in northern Israel, ...

[Etymology](#) - [Early history](#) - [Modern history](#) - [Demographics](#)

en.wikipedia.org/wiki/Haifa - [Cached](#) - [Similar](#)

[Haifa Wehbe](#) - [Wikipedia, the free encyclopedia](#) ☆

Haifa Wehbe (Arabic: هيفاء وهبي) is a Lebanese model, actress, and singer who rose to fame in the Arab world as runner up for Miss Lebanon and later ...

en.wikipedia.org/wiki/Haifa_Wehbe - 9 hours ago - [Cached](#) - [Similar](#)

[Dupont, The official Google Blog, 2010] "*Stars in search replace SearchWiki*"
<http://googleblog.blogspot.com/2010/03/stars-make-search-more-personal.html>



Stars in Search disappeared Your "starred" results remain



haifa

Search

About 7,310,000 results (0.31 seconds)

[Advanced search](#)

Everything

Maps

Images

Videos

More

Any time

Past 2 days

All results

Sites with images

More search tools

Something different

[tel aviv](#)

[netanya](#)

[ashdod](#)

[ashkelon](#)

[hadera](#)

Hotels in Haifa

www.booking.com/haifa-Hotels Book your hotel in **Haifa** online No reservation costs. Great rates!

Sponsored link



[Haifa](#)
[Israel](#)

maps.google.com

[Hotels](#) - [Restaurants](#) - [Bahai Gardens](#) - [Akko](#) - [Caesarea](#) - [Haifa Port](#) - [Nazareth](#) - [Technion](#)

Starred results for haifa

★ [Haifa - Wikipedia, the free encyclopedia](http://en.wikipedia.org/wiki/Haifa) - en.wikipedia.org/wiki/Haifa

★ [Haifa travel and tourism city guide and hotels in haifa - inisrael.com](http://www.inisrael.com/tour/haifa/index.html) - www.inisrael.com/tour/haifa/index.html

★ [Computer Science Department, Technion](http://www.cs.technion.ac.il/) - www.cs.technion.ac.il/

Haifa - Wikipedia, the free encyclopedia ☆

Haifa (Hebrew: [הַיְפָא](#) About this sound [Heifa](#) (help·info); Arabic: [حَيْفَا](#) About this sound [Hayfā](#) (help·info)) is the largest city in northern Israel, ...

[Etymology](#) - [Early history](#) - [Modern history](#) - [Demographics](#)
en.wikipedia.org/wiki/Haifa - [Cached](#) - [Similar](#)

Haifa Wehbe - Wikipedia, the free encyclopedia ☆

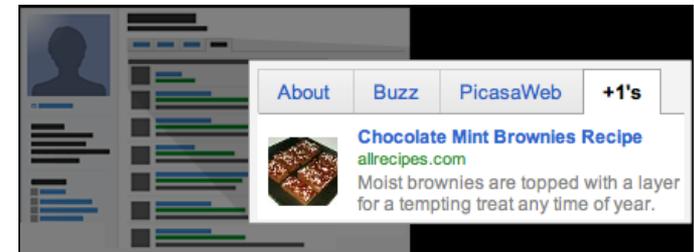
Haifa Wehbe (Arabic: [هيفاء وهبي](#)) is a Lebanese model, actress, and singer who rose to fame in the Arab world as runner up for Miss Lebanon and later ...

en.wikipedia.org/wiki/Haifa_Wehbe - [Cached](#) - [Similar](#)



From *personal* stars to *public* Google +1

- Results annotated with +1 button
- “Click +1 to publicly give something your stamp of approval”
- Visible in the search results of your contacts by default

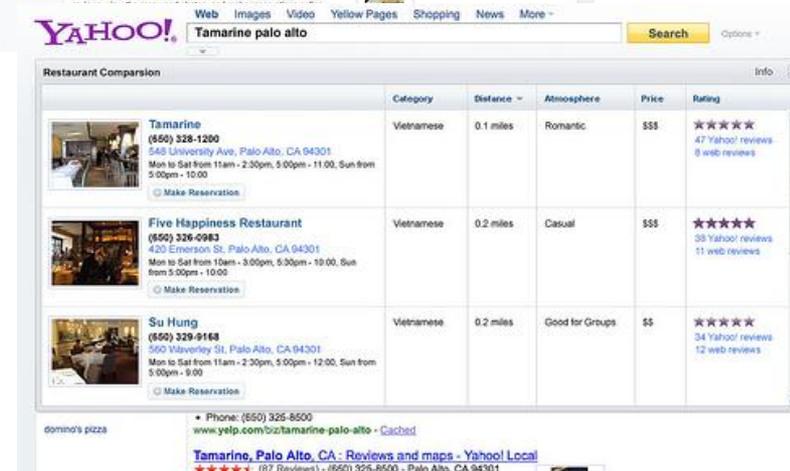


See <http://www.google.com/+1/button/>



Yahoo! QuickApps

- QuickApps displays applications relative to the word or phrase that you searched on in the search box
- Examples
 - Sketch a search for query "restaurants 94109"
 - OpenTable for "evvia palo alto"
 - Restaurant Comparison for "tamarine palo alto"





Beyond explicit search – explicit introspection

- Two examples
 - Google trends
 - Yahoo! Clues



Query trends exposed to users

- Time series of search trends, based on “query shares” of query term q at time t_i in geo location geo_j where

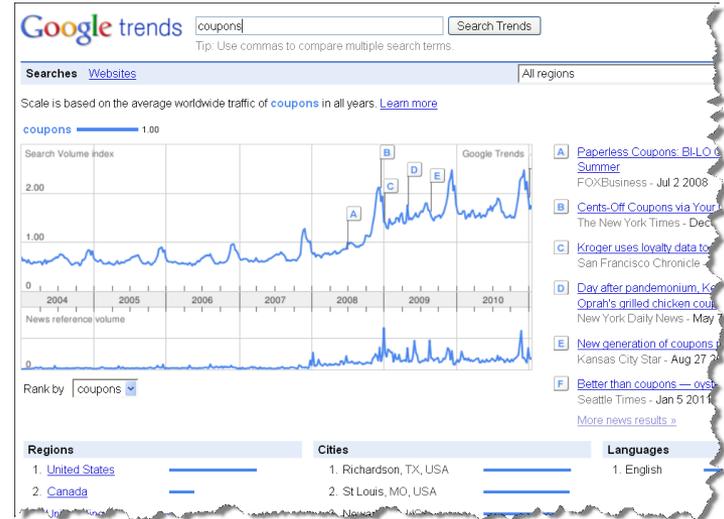
$$\text{query share} = \left(\frac{\text{number of queries for } q}{\text{number of queries}} \right) (t_i, geo_j)$$

- Google trends

- “Google Trends provides an index of the volume of Google queries by geographic location and category”
- [Varian and Choi 2009 -on Google Search Blog] “Predicting the Present with Google Trends”

- Google Insight for Search

- Uses the same data but geared to researchers and advertisers.
- [Shimshoni et al. 2009 on Google Search Blog] “On the Predictability of Search Trends ”





Yahoo! Clues <http://clues.yahoo.com>

Consider other facets

- Searches over time
- **By demographic**
- By Location
- **Search Flow**
- Related Search

angelina jolie Collapse

SEARCHES OVER TIME Describe these numbers

BY DEMOGRAPHIC Age Gender Both

Women	Age	Men
7%	Below 24	7%
10%	25 to 34	12%
9%	35 to 44	13%
7%	45 to 54	13%
4%	55 to 64	10%
2%	Above 65	6%

BY LOCATION

World >

1. United States	100	6. Malaysia	72
2. Iran	79	7. India	72
3. United Kingdom	76	8. France	71
4. Philippines	76	9. Indonesia	70
5. Canada	74	10. Brazil	66

SEARCH FLOW

Previous Queries	Next Queries
angelina+jolie angel angelina angelica pangniban angelina jolie movies	angelina jolie craigslist los angeles angelina jolie baby angelina jolie jennifer aniston angelina jolie hot brad pitt angelina jolie

RELATED SEARCHES

megan fox, britney spears, kate middleton, jennifer lopez, jennifer aniston, salma hayek, lindsay lohan, anne hathaway, kim kardashian, jennifer garner



By demographic

Combine Yahoo! search query log with:

- profile information provided by Yahoo! 28 M user million of its users (birth year, gender, ZIP code)
- US-census information aggregated by ZIP code

Annotate each query with ZIP code area



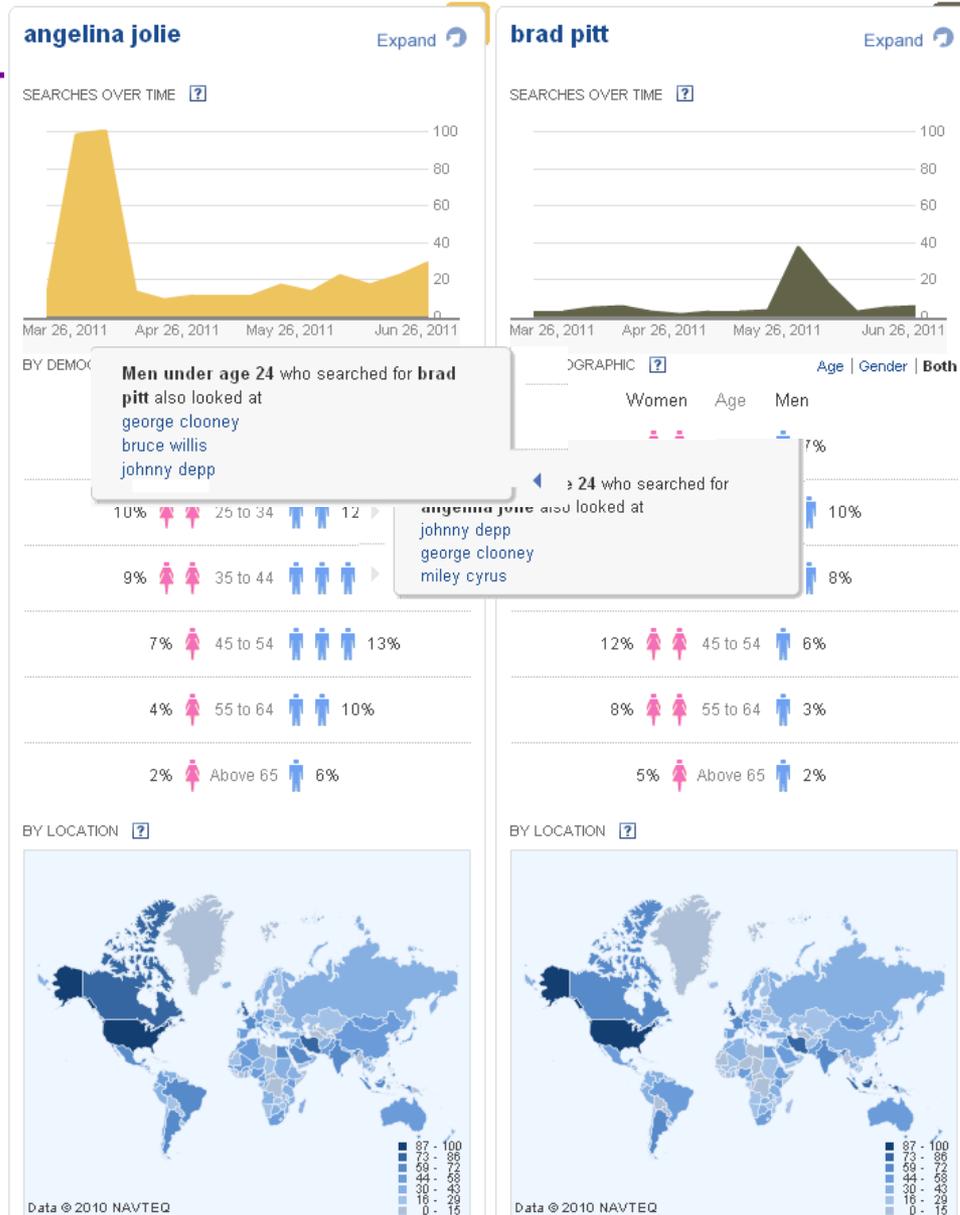
[Weber & Castillo SIGIR'2010]
"The Demographics of Web Search"
 [Weber & Jaimes, WSDM'2011]
"Who Uses Web Search for What? And How?"

location based models from geocoded queries
 [Serdyukov, Murdock & van Zwol SIGIR'2009]
"Placing Flickr Photos on a Map"





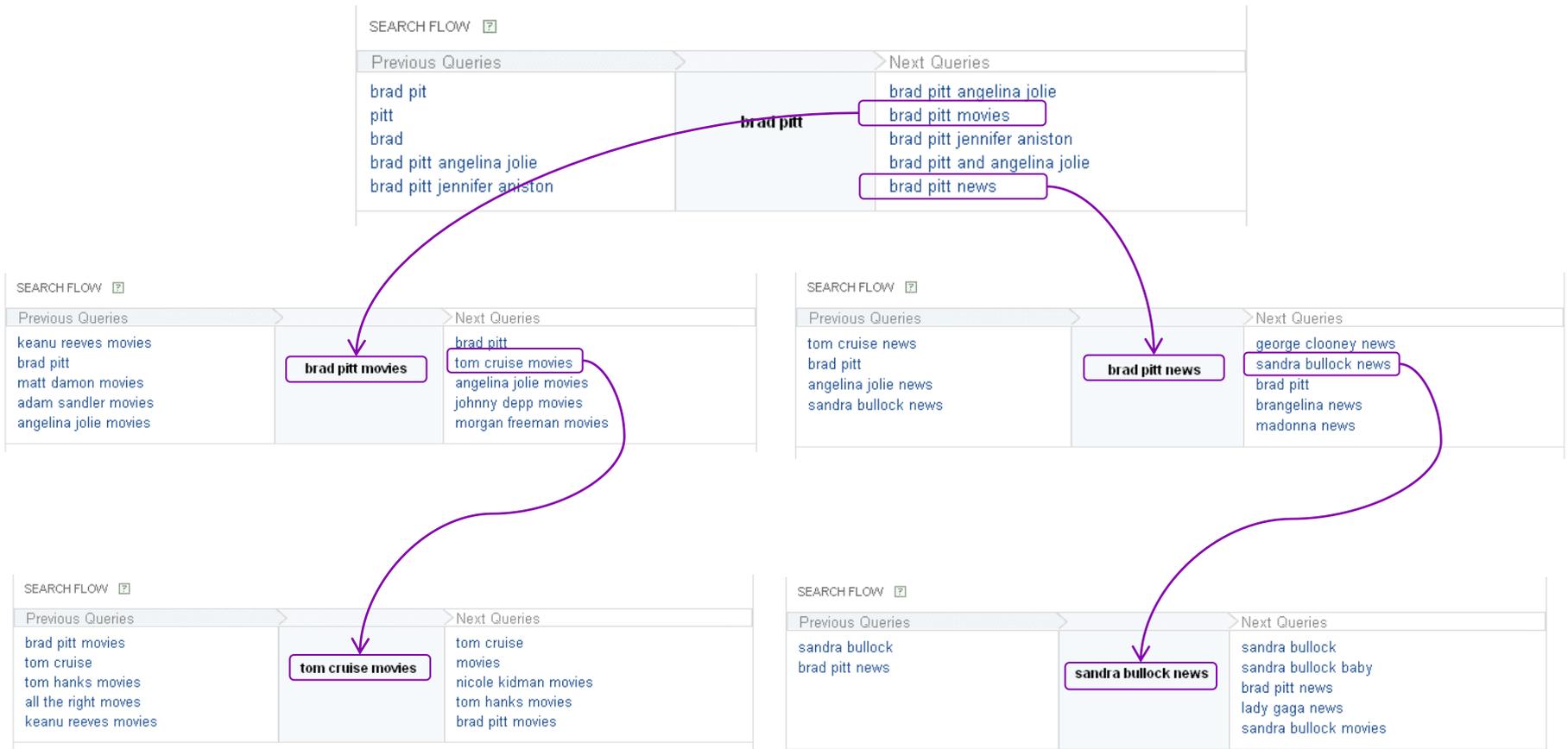
Comparing demographics on 2 queries





Query graph flow

- From implicit to explicit – showing the search flow to users
[Boldi et al, CIKM 2008] *"The query-flow graph: model and applications"*





Top Trends

Reset All ?

Time

< Jun 3, 2011 Today >

Number of months ▲ ▼

Day Week **Month** Year

Gender



Age

All

Below 24

25 to 34

35 to 44

45 to 54

55 to 64

Above 65

Location



Data © 2010 NAVTEQ

Zoom Out

Search by location

Category

All

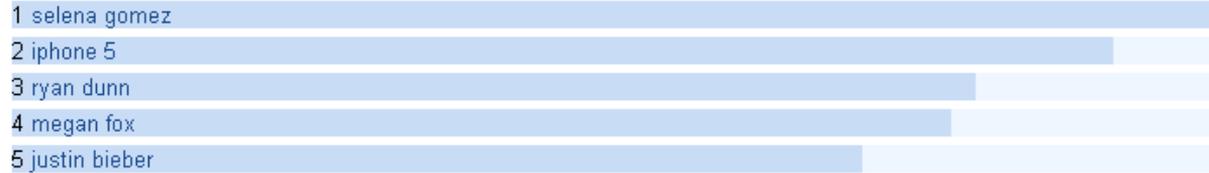
Sports

People

Places

Most Popular Search Queries

Filters Jun 3, 2011 to Today Below 24 x



1 selena gomez

Will **Selena Gomez** be the next Charlotte York in 'S...', TheCelebrityCafe.com
Selena Gomez And Leighton Meester Talk 'Monte Carl...', omg!
Selena Gomez: 'Rock of Ages' Gall, Just Jared Jr.



2 iphone 5

Link: The iPhone Turns Four: How It Has Changed Us..., Forbes
 iPhone birthday: How Apple forever changed the sma..., The Christian Science Monitor
iPhone 5, iPad 3 launching this October?, ZDNet

3 ryan dunn

Ryan Dunn honored by 'Jackass' crew with a tribute..., KABC-TV Los Angeles
Ryan Dunn honored with a tribute video - Watch, KABC-TV Los Angeles
Ryan Dunn Death Site Vandalized Recap, antiMUSIC



4 megan fox

Megan Fox's rep confirms Shia LaBeouf hookup, TheCelebrityCafe.com
 Rep confirms **Megan Fox**, Shia LaBeouf fling, TheCelebrityCafe.com
 Fox's rep confirms onset tryst with LaBeouf, Hollywood.com



5 justin bieber

Justin Bieber Vs 'Douche Bag' Recap, antiMUSIC
Justin Bieber's diva-sized ego tempered by Selena ..., Chicago Sun-Times
Justin Bieber Gets Inspirational In Google Chrome ..., Billboard



Conclusion and future directions

Session 4



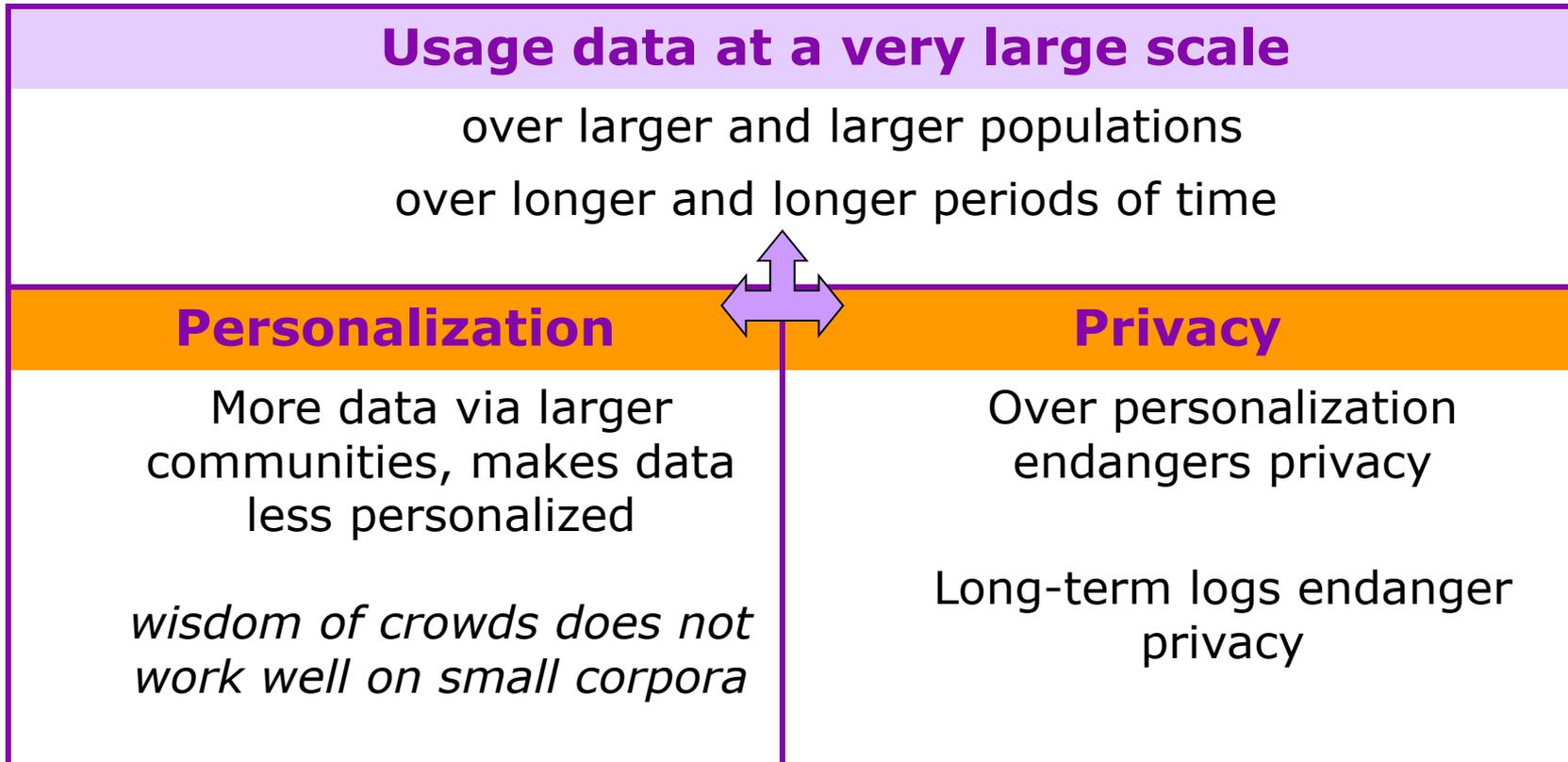


Conclusions

- We still don't understand well information needs
 - Will we ever?
 - Queries are just an approximation and brain electrodes won't work☺
- Users crave for
 - more interactive features
 - more "digest" features – show me answers not links
 - more personalization but are scared of privacy infringements



We are far from being done with innovation in search engines
Large scale usage data is key BUT





- Front-end and user experience
 - The most probable reason for users to switch between quasi-equivalent engines is a better user experience
- Depart from the rectangle/ranked list paradigm
 - Get rid of queries? **Implicit search**
 - Content delivery is one flavor
 - But in general, why should we even have to formulate a query?