Use of Click Data for Web Search

Tao Yang

UCSB 290N

Table of Content

- Search Engine Logs
- Eyetracking data on position bias
  - Click data for ranker training [Joachims, KDD02]
- Case study: Use of click data for search ranking [Agichtein et al, SIGIR 06]

Search Logs

- Query logs recorded by search engines

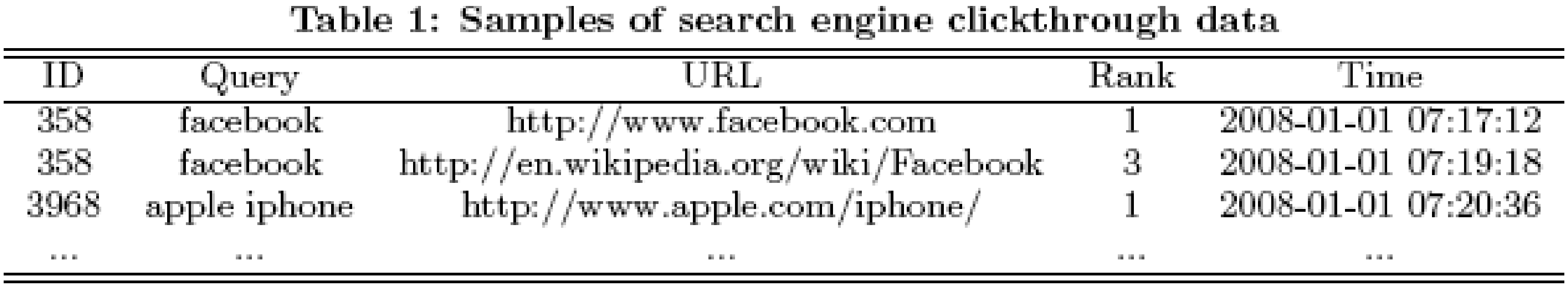
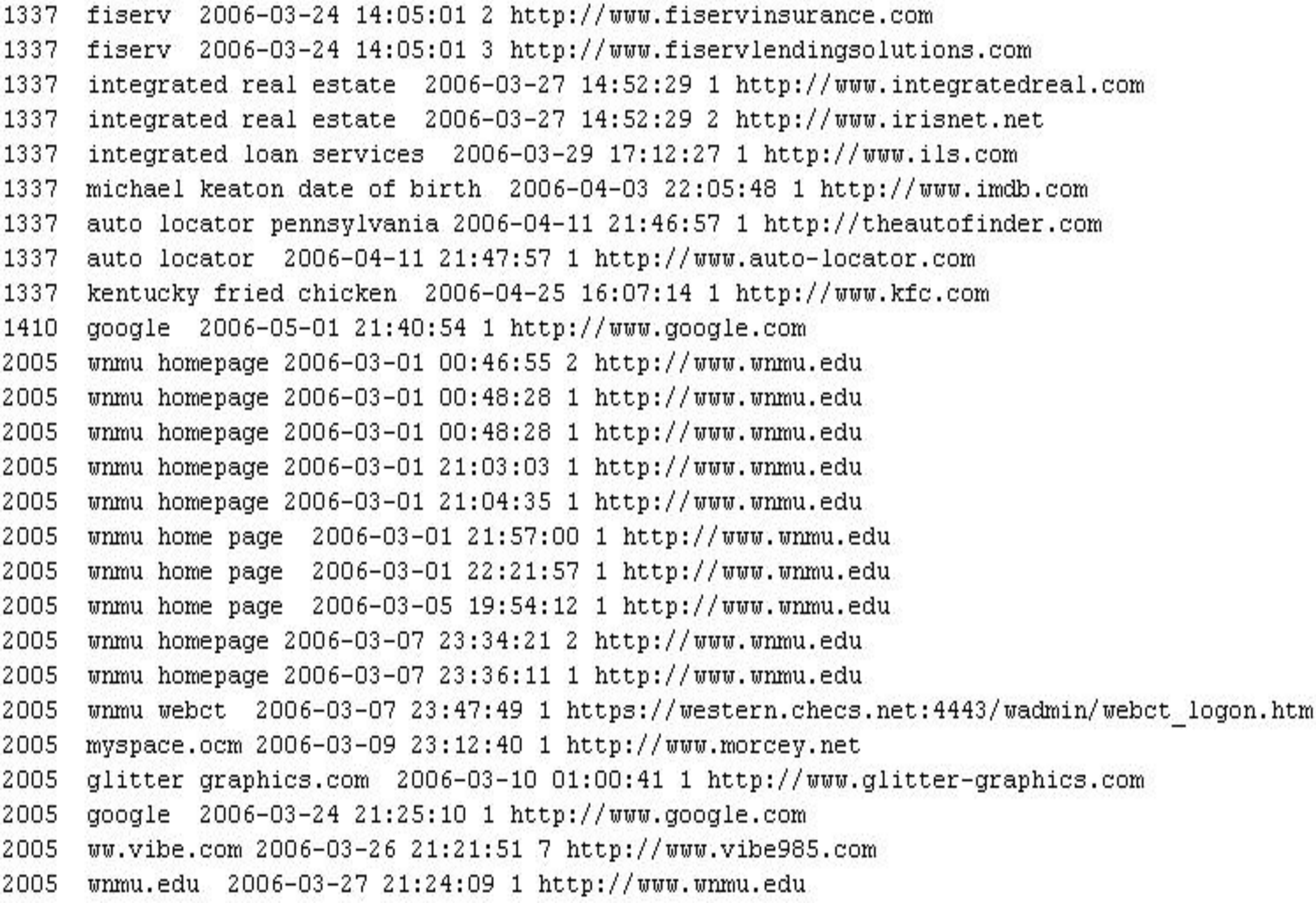


Table 1: Samples of search engine clickthrough data

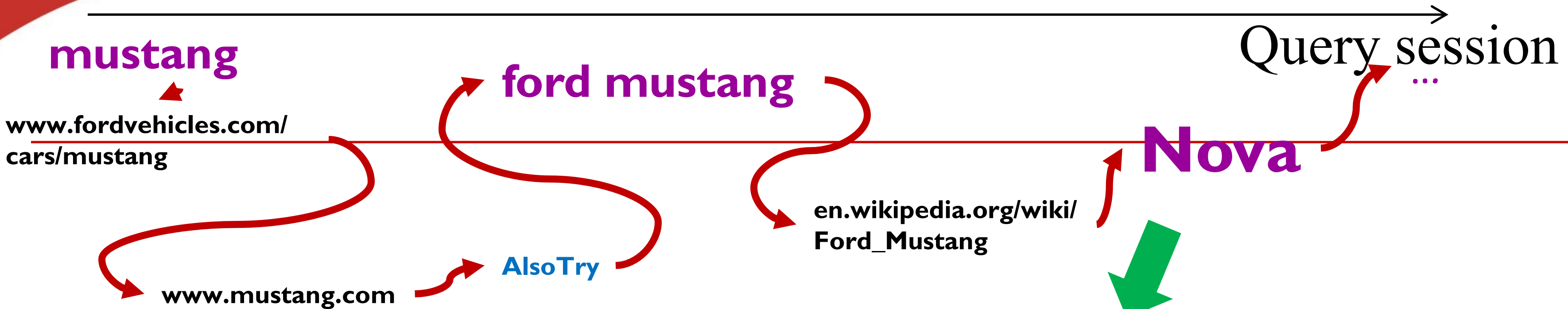
ID	Query	URL	Rank	Time
358	facebook	http://www.facebook.com	1	2008-01-01 07:17:12
358	facebook	http://en.wikipedia.org/wiki/Facebook	3	2008-01-01 07:19:18
3968	apple iphone	http://www.apple.com/iphone/	1	2008-01-01 07:20:36
...	...	...	...	...

- Huge amount of data: e.g. 10TB/day at Bing



```
1337  fiserv  2006-03-24 14:05:01 2 http://www.fiservinsurance.com
1337  fiserv  2006-03-24 14:05:01 3 http://www.fiservlendingsolutions.com
1337  integrated real estate  2006-03-27 14:52:29 1 http://www.integratedreal.com
1337  integrated real estate  2006-03-27 14:52:29 2 http://www.irisnet.net
1337  integrated loan services  2006-03-29 17:12:27 1 http://www.ils.com
1337  michael keaton date of birth  2006-04-03 22:05:48 1 http://www.imdb.com
1337  auto locator pennsylvania 2006-04-11 21:46:57 1 http://theautofinder.com
1337  auto locator  2006-04-11 21:47:57 1 http://www.auto-locator.com
1337  kentucky fried chicken  2006-04-25 16:07:14 1 http://www.kfc.com
1410  google  2006-05-01 21:40:54 1 http://www.google.com
2005  wnmu homepage 2006-03-01 00:46:55 2 http://www.wnmu.edu
2005  wnmu homepage 2006-03-01 00:48:28 1 http://www.wnmu.edu
2005  wnmu homepage 2006-03-01 00:48:28 1 http://www.wnmu.edu
2005  wnmu homepage 2006-03-01 21:03:03 1 http://www.wnmu.edu
2005  wnmu homepage 2006-03-01 21:04:35 1 http://www.wnmu.edu
2005  wnmu home page  2006-03-01 21:57:00 1 http://www.wnmu.edu
2005  wnmu home page  2006-03-01 22:21:57 1 http://www.wnmu.edu
2005  wnmu home page  2006-03-05 19:54:12 1 http://www.wnmu.edu
2005  wnmu homepage 2006-03-07 23:34:21 2 http://www.wnmu.edu
2005  wnmu homepage 2006-03-07 23:36:11 1 http://www.wnmu.edu
2005  wnmu webct  2006-03-07 23:47:49 1 https://western.checs.net:4443/wadmin/webct_logon.htm
2005  myspace.ocm 2006-03-09 23:12:40 1 http://www.morcey.net
2005  glitter graphics.com  2006-03-10 01:00:41 1 http://www.glitter-graphics.com
2005  google  2006-03-24 21:25:10 1 http://www.google.com
2005  ww.vibe.com 2006-03-26 21:21:51 7 http://www.vibe985.com
2005  wnmu.edu  2006-03-27 21:24:09 1 http://www.wnmu.edu
```

Search sessions



Query session …

mustang

www.fordvehicles.com/cars/mustang

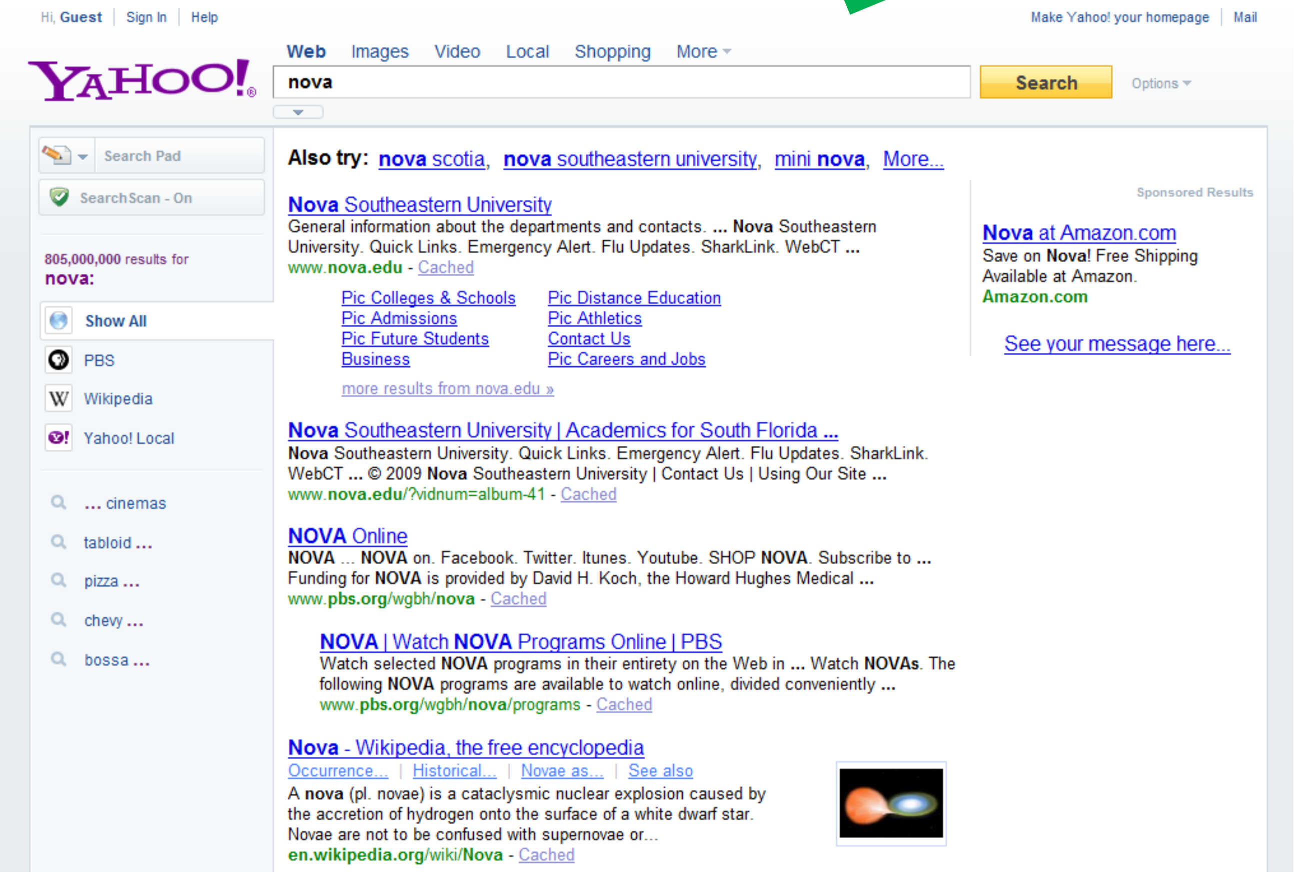
www.mustang.com

AlsoTry

ford mustang

en.wikipedia.org/wiki/Ford_Mustang

Nova



Hi, Guest | Sign In | Help
Make Yahoo! your homepage | Mail
YAHOO!
Web Images Video Local Shopping More
Search
Options
Search Pad
SearchScan - On

nova

805,000,000 results for **nova:**

Also try: nova scotia, nova southeastern university, mini nova, More...

Nova Southeastern University
General information about the departments and contacts. ... **Nova** Southeastern University. Quick Links. Emergency Alert. Flu Updates. SharkLink. WebCT ...
www.**nova**.edu - Cached

Pic Colleges & Schools | Pic Distance Education
Pic Admissions | Pic Athletics
Pic Future Students | Contact Us
Business | Pic Careers and Jobs

more results from nova.edu »

Nova Southeastern University | Academics for South Florida ...
Nova Southeastern University. Quick Links. Emergency Alert. Flu Updates. SharkLink. WebCT ... © 2009 **Nova** Southeastern University | Contact Us | Using Our Site ...
www.**nova**.edu/?vidnum=album-41 - Cached

NOVA Online
NOVA ... **NOVA** on. Facebook. Twitter. Itunes. Youtube. SHOP **NOVA**. Subscribe to ... Funding for **NOVA** is provided by David H. Koch, the Howard Hughes Medical ...
www.**pbs.org**/wgbh/**nova** - Cached

NOVA | Watch NOVA Programs Online | PBS
Watch selected **NOVA** programs in their entirety on the Web in ... Watch **NOVAs**. The following **NOVA** programs are available to watch online, divided conveniently ...
www.**pbs.org**/wgbh/**nova**/programs - Cached

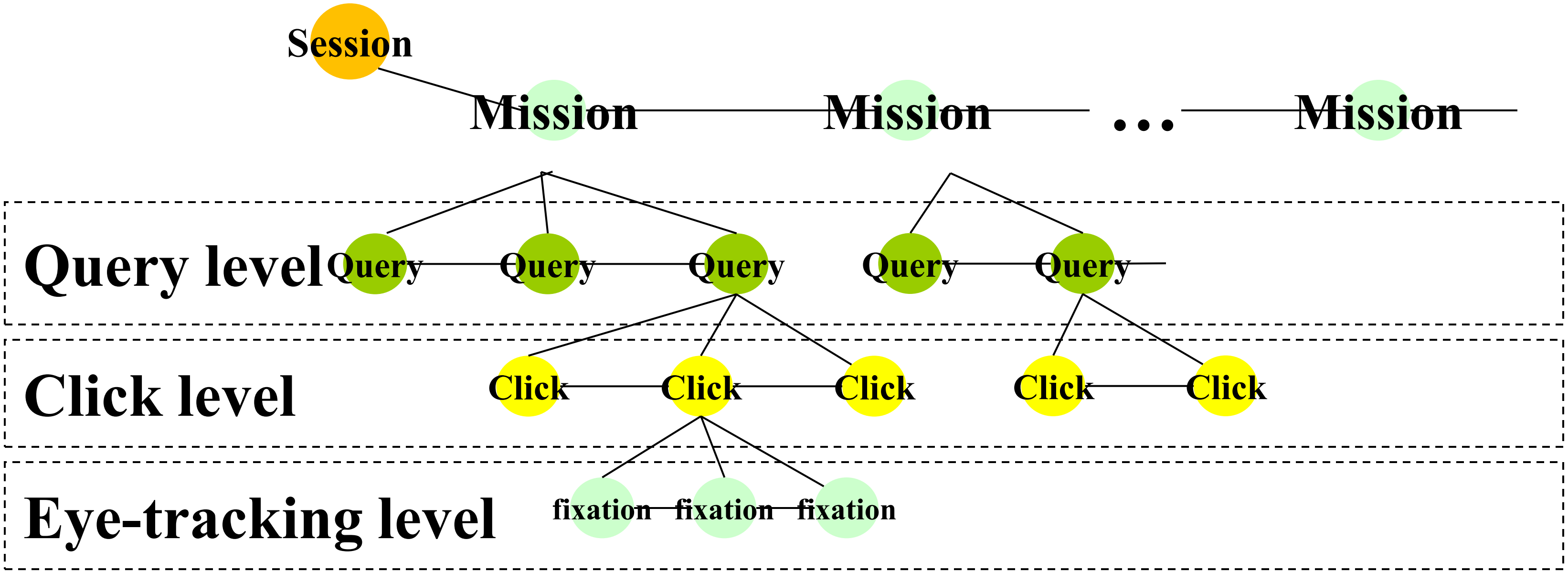
Nova - Wikipedia, the free encyclopedia
Occurrence... | Historical... | Novae as... | See also
A **nova** (pl. novae) is a cataclysmic nuclear explosion caused by the accretion of hydrogen onto the surface of a white dwarf star. Novae are not to be confused with supernovae or...
en.wikipedia.org/wiki/**Nova** - Cached

Sponsored Results

Nova at Amazon.com
Save on **Nova**! Free Shipping Available at Amazon.
Amazon.com

See your message here...

Query sessions and analysis



Query-URL correlations:

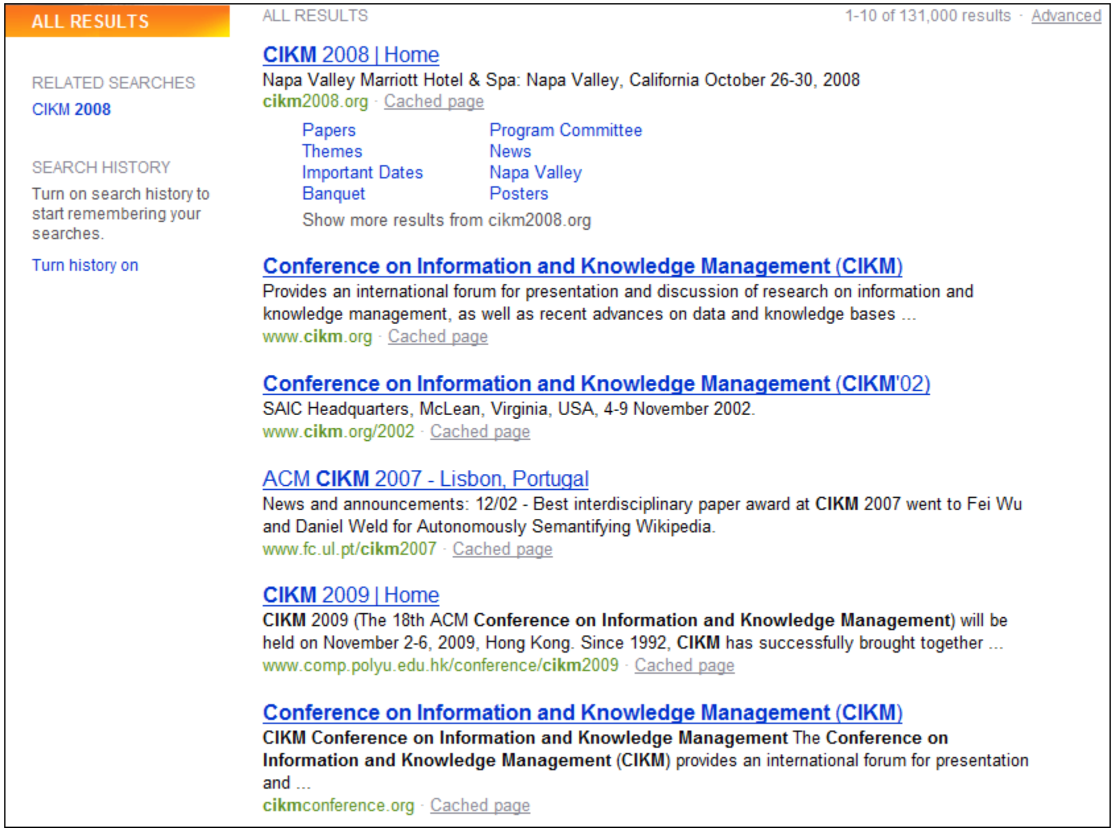
- Query-to-pick
- Query-to-query
- Pick-to-pick

Examples of behavior analysis with search logs

- **Query-pick (click) analysis**
- **Session detection**
- **Classification**
  - $x_1, x_2, \ldots, x_N \rightarrow y$
  - eg, whether the session has a commercial intent
- **Sequence labeling**
  - $x_1, x_2, \ldots, x_N \rightarrow y_1, y_2, \ldots, y_N$
  - eg, segment a search sequence into missions and goals
- **Prediction**
  - $x_1, x_2, \ldots, x_{N-1} \rightarrow y_N$
- **Similarity**
  - Similarity(S_1, S_2)

Query-pick (click) analysis

- **Search Results for “CIKM”**



ALL RESULTS

RELATED SEARCHES

CIKM **2008**

SEARCH HISTORY

Turn on search history to start remembering your searches.

Turn history on

ALL RESULTS 1-10 of 131,000 results · Advanced

CIKM 2008 | Home
Napa Valley Marriott Hotel & Spa: Napa Valley, California October 26-30, 2008
cikm2008.org · Cached page

Papers	Program Committee
Themes	News
Important Dates	Napa Valley
Banquet	Posters

Show more results from cikm2008.org

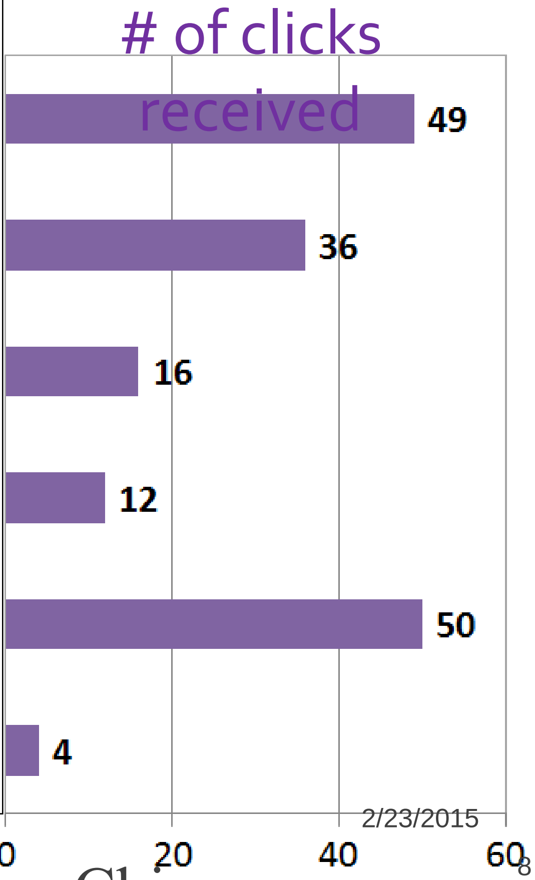
Conference on Information and Knowledge Management (CIKM)
Provides an international forum for presentation and discussion of research on information and knowledge management, as well as recent advances on data and knowledge bases ...
www.**cikm**.org · Cached page

Conference on Information and Knowledge Management (CIKM'02)
SAIC Headquarters, McLean, Virginia, USA, 4-9 November 2002.
www.**cikm**.org/2002 · Cached page

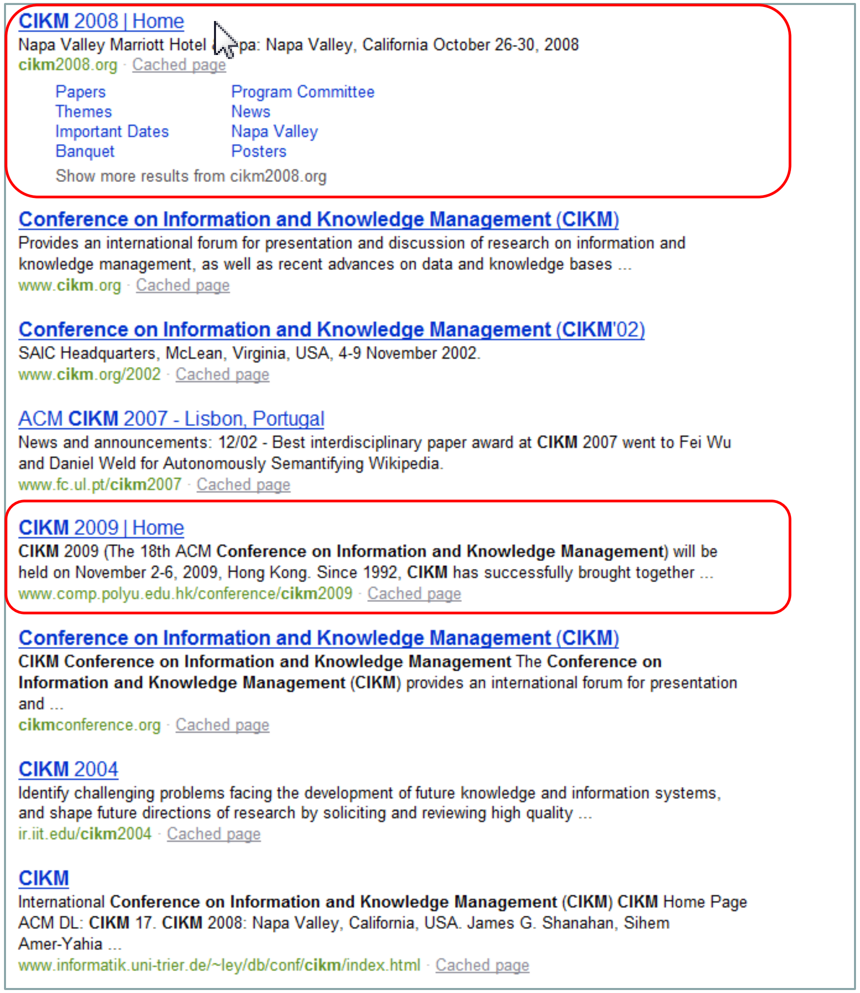
ACM **CIKM** 2007 - Lisbon, Portugal
News and announcements: 12/02 - Best interdisciplinary paper award at **CIKM** 2007 went to Fei Wu and Daniel Weld for Autonomously Semantifying Wikipedia.
www.fc.ul.pt/**cikm**2007 · Cached page

CIKM 2009 | Home
CIKM 2009 (The 18th ACM **Conference on Information and Knowledge Management**) will be held on November 2-6, 2009, Hong Kong. Since 1992, **CIKM** has successfully brought together ...
www.comp.polyu.edu.hk/conference/**cikm**2009 · Cached page

Conference on Information and Knowledge Management (CIKM)
CIKM Conference on Information and Knowledge Management The **Conference on Information and Knowledge Management** (**CIKM**) provides an international forum for presentation and ...
cikmconference.org · Cached page



Interpret Clicks: an Example



[CIKM 2008 | Home](#)
Napa Valley Marriott Hotel & Spa: Napa Valley, California October 26-30, 2008
cikm2008.org · Cached page

Papers · Program Committee
Themes · News
Important Dates · Napa Valley
Banquet · Posters
Show more results from cikm2008.org

Conference on Information and Knowledge Management (CIKM)
Provides an international forum for presentation and discussion of research on information and knowledge management, as well as recent advances on data and knowledge bases ...
www.cikm.org · Cached page

Conference on Information and Knowledge Management (CIKM'02)
SAIC Headquarters, McLean, Virginia, USA, 4-9 November 2002.
www.cikm.org/2002 · Cached page

ACM CIKM 2007 - Lisbon, Portugal
News and announcements: 12/02 - Best interdisciplinary paper award at CIKM 2007 went to Fei Wu and Daniel Weld for Autonomously Semantifying Wikipedia.
www.fc.ul.pt/cikm2007 · Cached page

CIKM 2009 | Home
CIKM 2009 (The 18th ACM Conference on Information and Knowledge Management) will be held on November 2-6, 2009, Hong Kong. Since 1992, CIKM has successfully brought together ...
www.comp.polyu.edu.hk/conference/cikm2009 · Cached page

Conference on Information and Knowledge Management (CIKM)
CIKM Conference on Information and Knowledge Management The Conference on Information and Knowledge Management (CIKM) provides an international forum for presentation and ...
cikmconference.org · Cached page

CIKM 2004
Identify challenging problems facing the development of future knowledge and information systems, and shape future directions of research by soliciting and reviewing high quality ...
ir.iit.edu/cikm2004 · Cached page

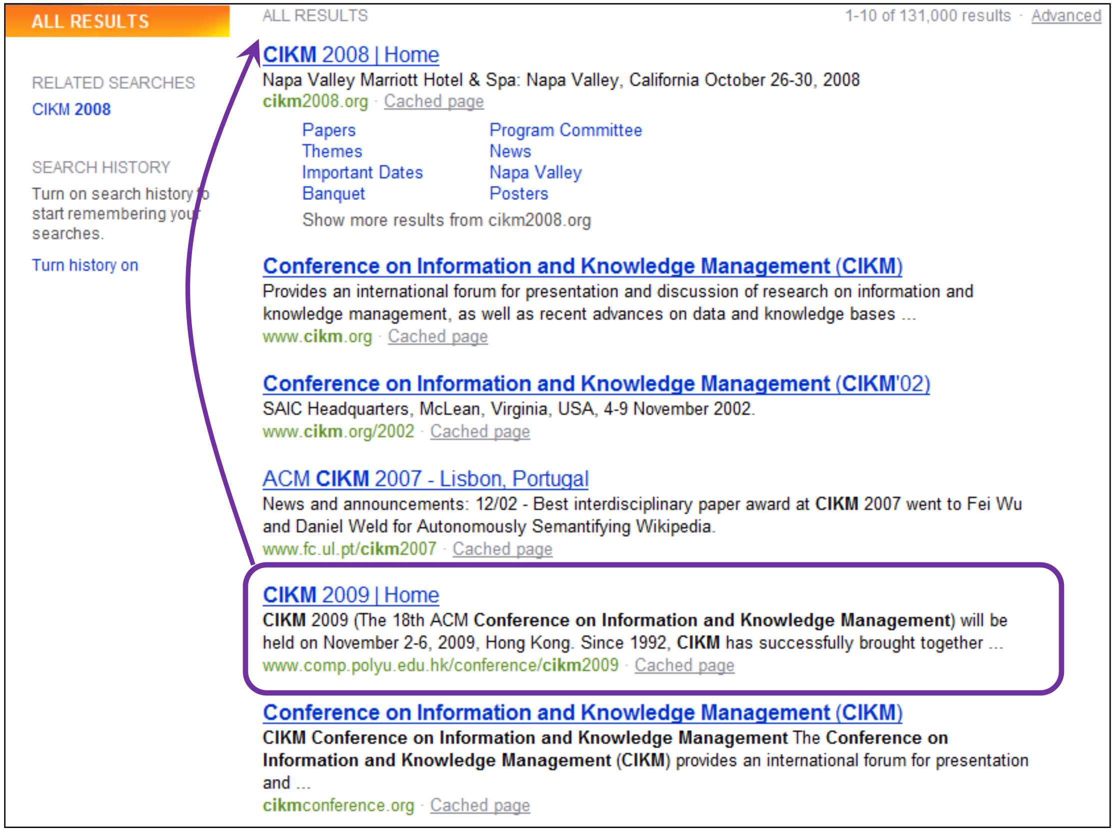
CIKM
International Conference on Information and Knowledge Management (CIKM) CIKM Home Page ACM DL: CIKM 17. CIKM 2008: Napa Valley, California, USA. James G. Shanahan, Sihem Amer-Yahia ...
www.informatik.uni-trier.de/~ley/db/conf/cikm/index.html · Cached page

- **Clicks are good…**
  - Are these two clicks equally “good”?
- **Non-clicks may have excuses:**
  - Not relevant
  - Not examined

2/23/2015

Use of behavior data

- **Adapt ranking to user clicks?**



ALL RESULTS

RELATED SEARCHES

CIKM **2008**

SEARCH HISTORY

Turn on search history to start remembering your searches.

Turn history on

ALL RESULTS 1-10 of 131,000 results · Advanced

CIKM 2008 | Home
Napa Valley Marriott Hotel & Spa: Napa Valley, California October 26-30, 2008
cikm2008.org · Cached page

Papers	Program Committee
Themes	News
Important Dates	Napa Valley
Banquet	Posters

Show more results from cikm2008.org

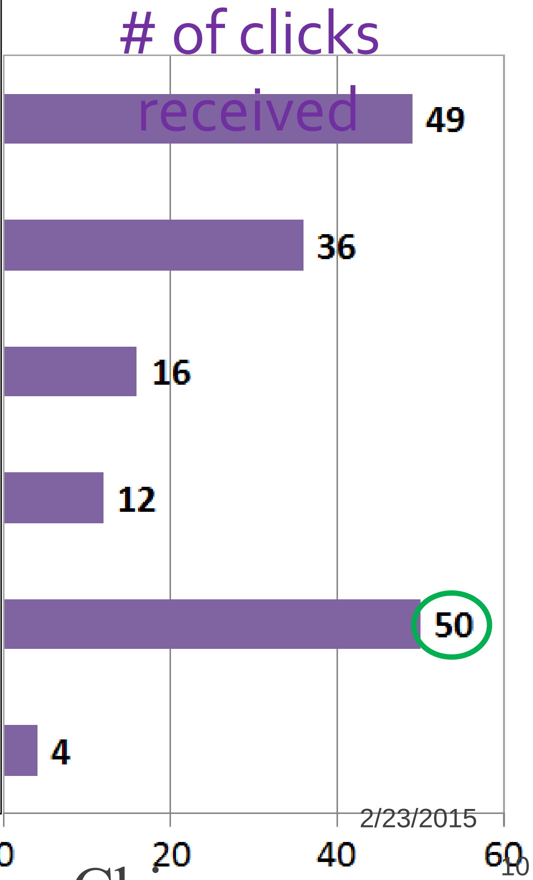
Conference on Information and Knowledge Management (CIKM)
Provides an international forum for presentation and discussion of research on information and knowledge management, as well as recent advances on data and knowledge bases ...
www.cikm.org · Cached page

Conference on Information and Knowledge Management (CIKM'02)
SAIC Headquarters, McLean, Virginia, USA, 4-9 November 2002.
www.cikm.org/2002 · Cached page

ACM CIKM 2007 - Lisbon, Portugal
News and announcements: 12/02 - Best interdisciplinary paper award at **CIKM** 2007 went to Fei Wu and Daniel Weld for Autonomously Semantifying Wikipedia.
www.fc.ul.pt/cikm2007 · Cached page

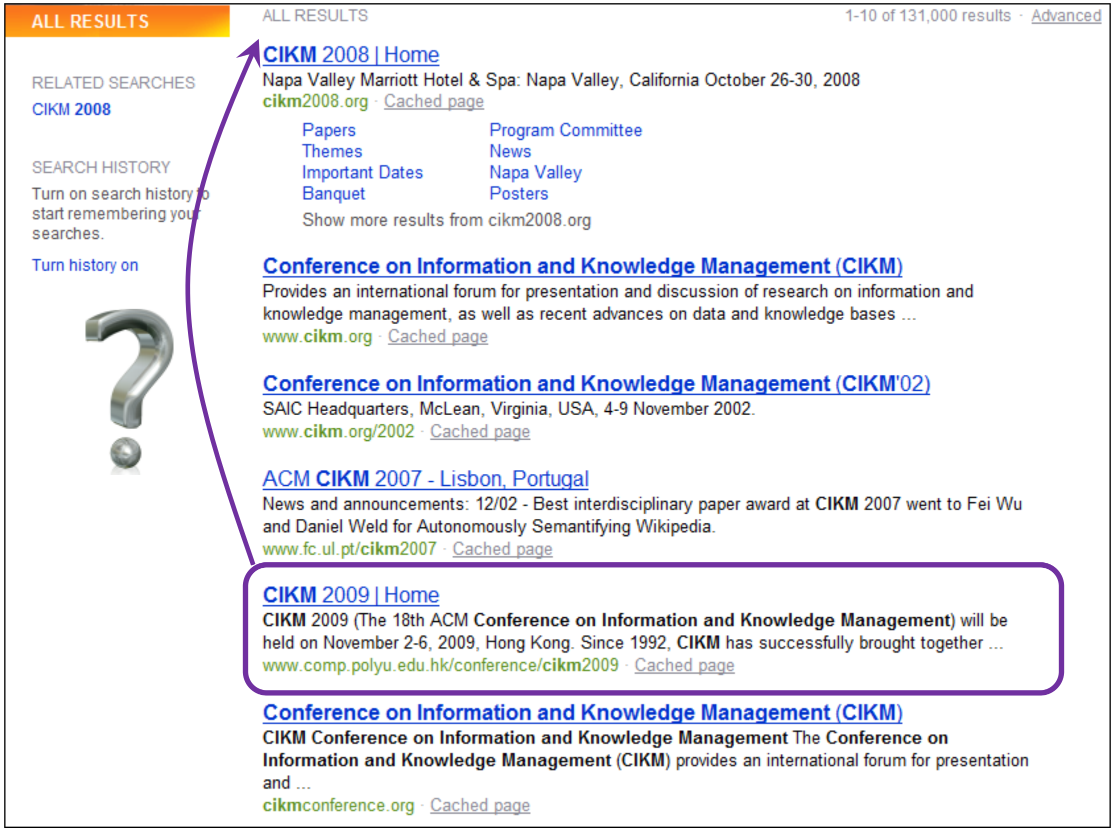
CIKM 2009 | Home
CIKM 2009 (The 18th ACM **Conference on Information and Knowledge Management**) will be held on November 2-6, 2009, Hong Kong. Since 1992, **CIKM** has successfully brought together ...
www.comp.polyu.edu.hk/conference/cikm2009 · Cached page

Conference on Information and Knowledge Management (CIKM)
CIKM Conference on Information and Knowledge Management The **Conference on Information and Knowledge Management** (**CIKM**) provides an international forum for presentation and ...
cikmconference.org · Cached page



Non-trivial cases

- **Tools needed for non-trivial cases**



Turn on search history to start remembering your searches.

1-10 of 131,000 results · Advanced

CIKM 2008 | Home
Napa Valley Marriott Hotel & Spa: Napa Valley, California October 26-30, 2008
cikm2008.org · Cached page

Papers	Program Committee
Themes	News
Important Dates	Napa Valley
Banquet	Posters

Show more results from cikm2008.org

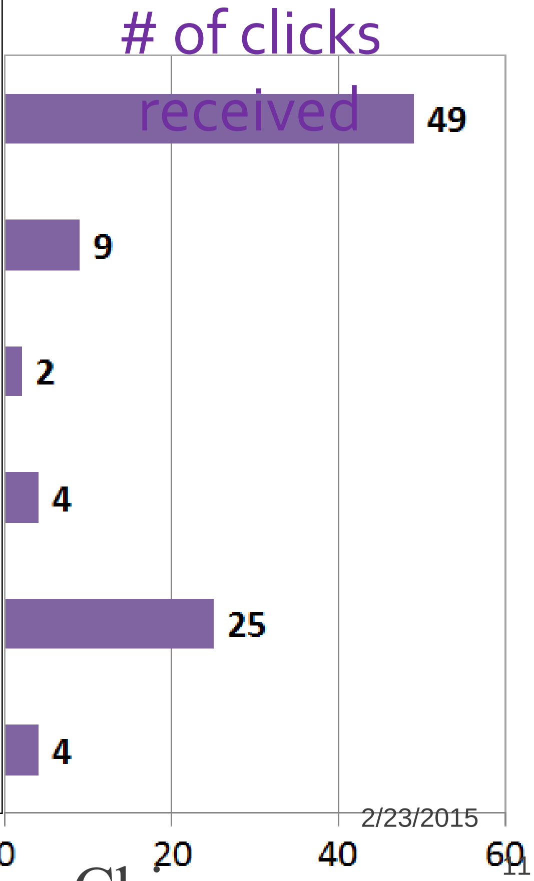
Conference on Information and Knowledge Management (CIKM)
Provides an international forum for presentation and discussion of research on information and knowledge management, as well as recent advances on data and knowledge bases ...
www.**cikm**.org · Cached page

Conference on Information and Knowledge Management (CIKM'02)
SAIC Headquarters, McLean, Virginia, USA, 4-9 November 2002.
www.**cikm**.org/2002 · Cached page

ACM **CIKM** 2007 - Lisbon, Portugal
News and announcements: 12/02 - Best interdisciplinary paper award at **CIKM** 2007 went to Fei Wu and Daniel Weld for Autonomously Semantifying Wikipedia.
www.fc.ul.pt/**cikm**2007 · Cached page

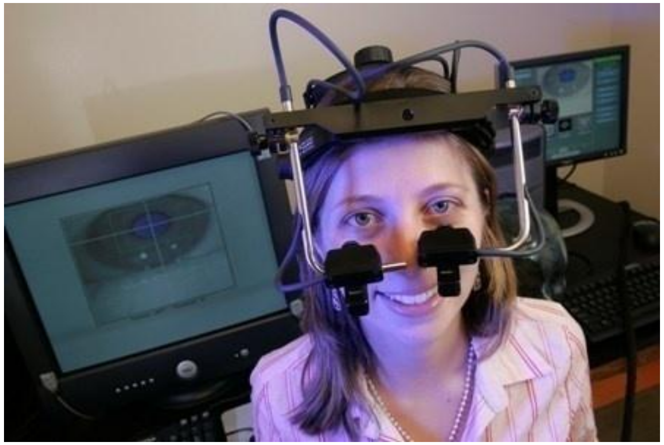
CIKM 2009 | Home
CIKM 2009 (The 18th ACM **Conference on Information and Knowledge Management**) will be held on November 2-6, 2009, Hong Kong. Since 1992, **CIKM** has successfully brought together ...
www.comp.polyu.edu.hk/conference/**cikm**2009 · Cached page

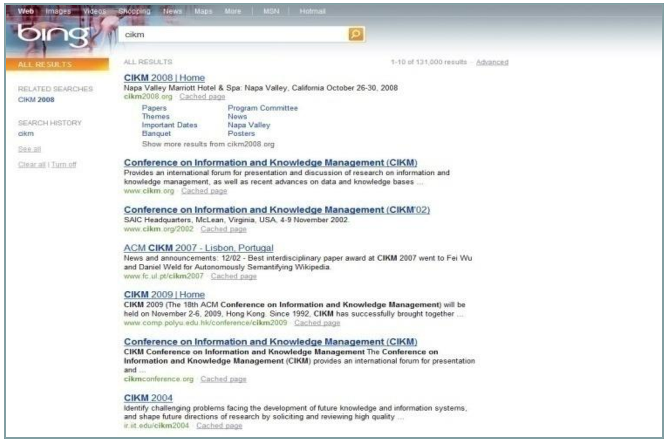
Conference on Information and Knowledge Management (CIKM)
CIKM Conference on Information and Knowledge Management The **Conference on Information and Knowledge Management** (**CIKM**) provides an international forum for presentation and ...
cikmconference.org · Cached page



2/23/2015

Eye-tracking User Study





cikm

ALL RESULTS 1-10 of 131,000 results · Advanced

CIKM 2008 | Home
Napa Valley Marriott Hotel & Spa: Napa Valley, California October 26-30, 2008
cikm2008.org · Cached page
Papers · Program Committee
Themes · News
Important Dates · Napa Valley
Banquet · Posters
Show more results from cikm2008.org

Conference on Information and Knowledge Management (CIKM)
Provides an international forum for presentation and discussion of research on information and knowledge management, as well as recent advances on data and knowledge bases …
www.cikm.org · Cached page

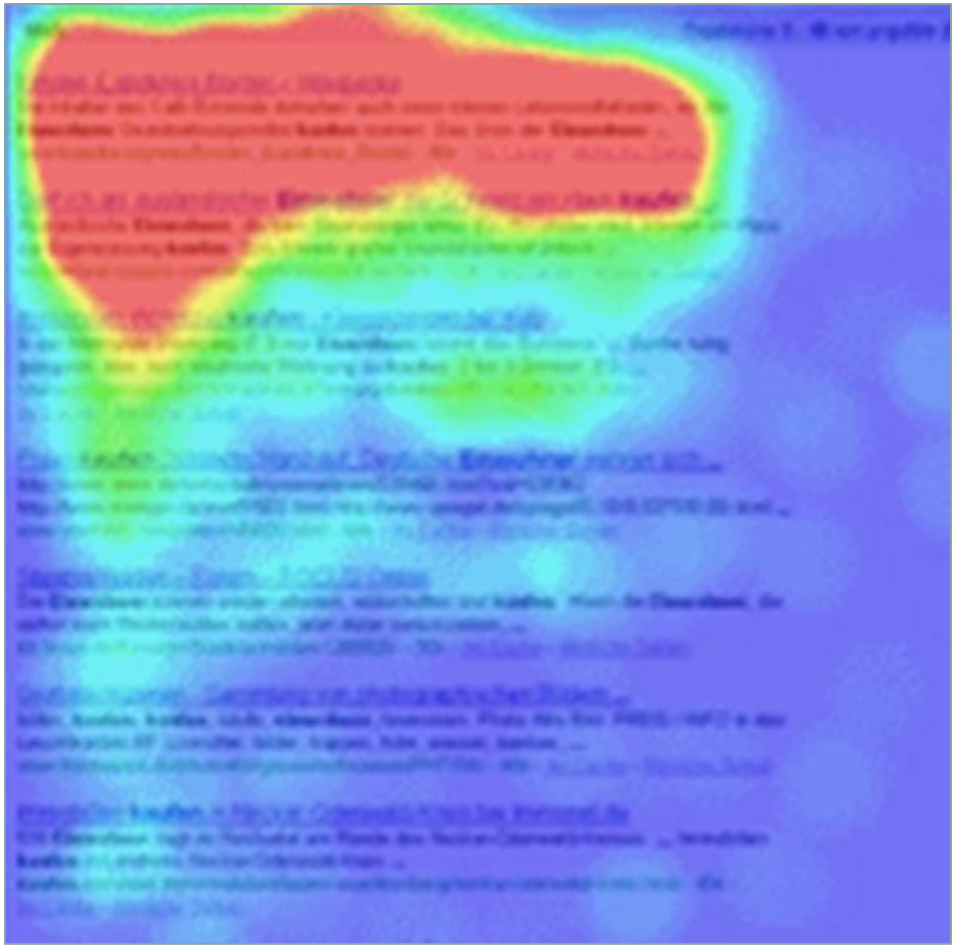
Conference on Information and Knowledge Management (CIKM'02)
SAIC Headquarters, McLean, Virginia, USA, 4-9 November 2002.
www.cikm.org/2002 · Cached page

ACM CIKM 2007 - Lisbon, Portugal
News and announcements: 12/02 - Best interdisciplinary paper award at CIKM 2007 went to Fei Wu and Daniel Weld for Autonomously Semantifying Wikipedia.
www.fc.ul.pt/cikm2007 · Cached page

CIKM 2009 | Home
CIKM 2009 (The 18th ACM Conference on Information and Knowledge Management) will be held on November 2-6, 2009, Hong Kong. Since 1992, CIKM has successfully brought together …
www.comp.polyu.edu.hk/conference/cikm2009 · Cached page

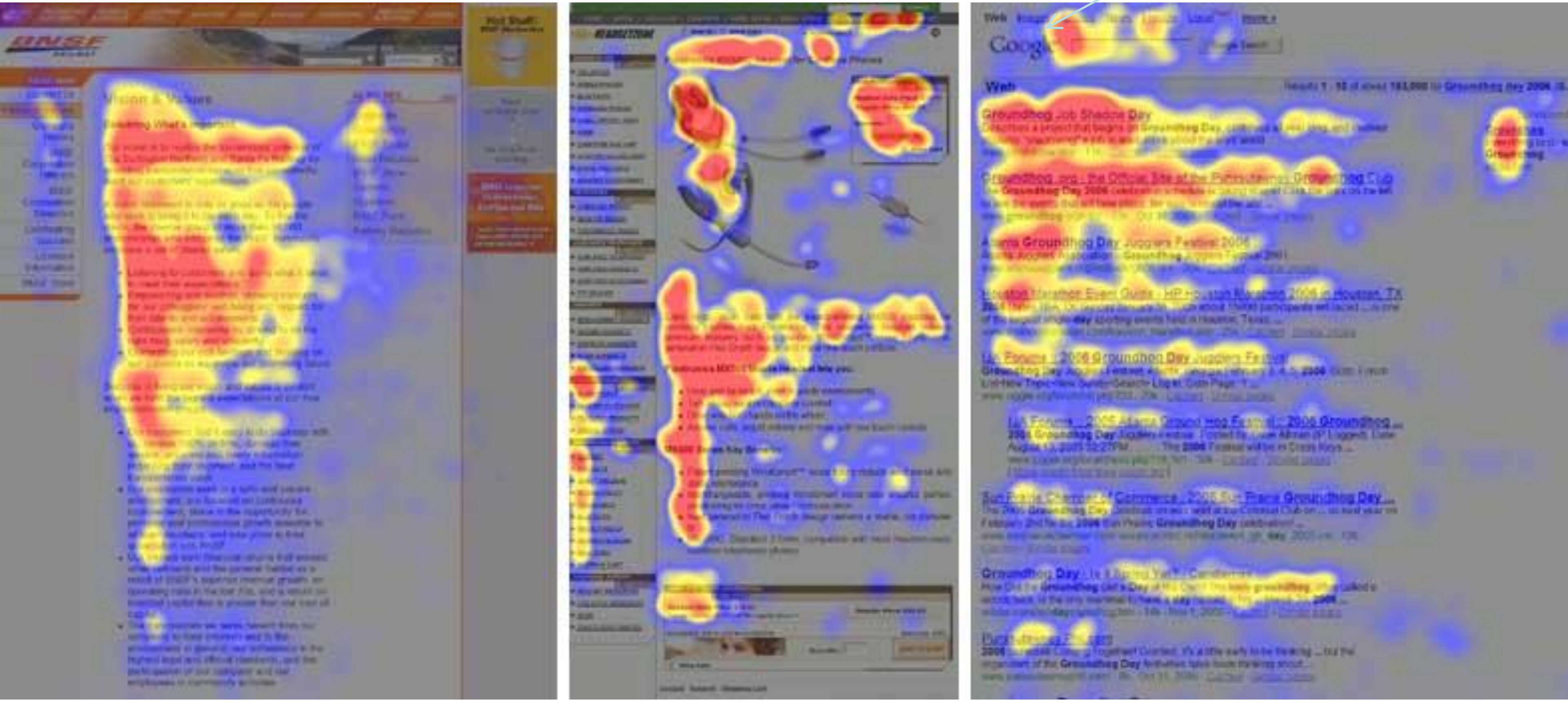
Conference on Information and Knowledge Management (CIKM)
CIKM Conference on Information and Knowledge Management The Conference on Information and Knowledge Management (CIKM) provides an international forum for presentation and …
cikmconference.org · Cached page

CIKM 2004
Identify challenging problems facing the development of future knowledge and information systems, and shape future directions of research by soliciting and reviewing high quality …
ir.iit.edu/cikm2004 · Cached page

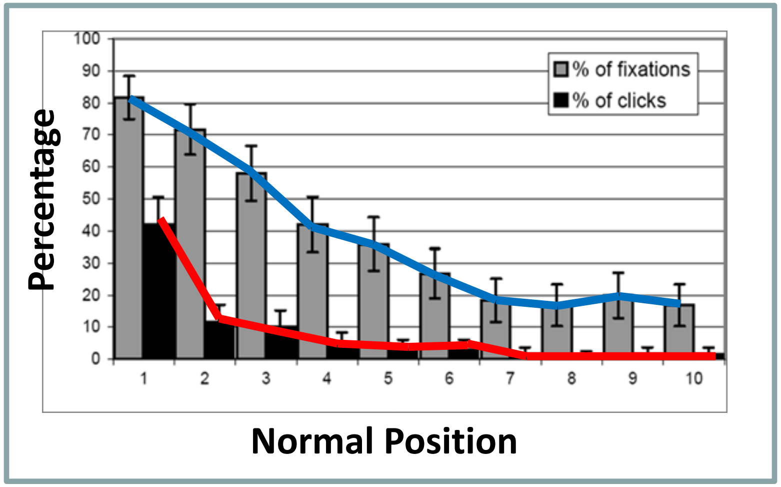


Eye tracking for different web sites

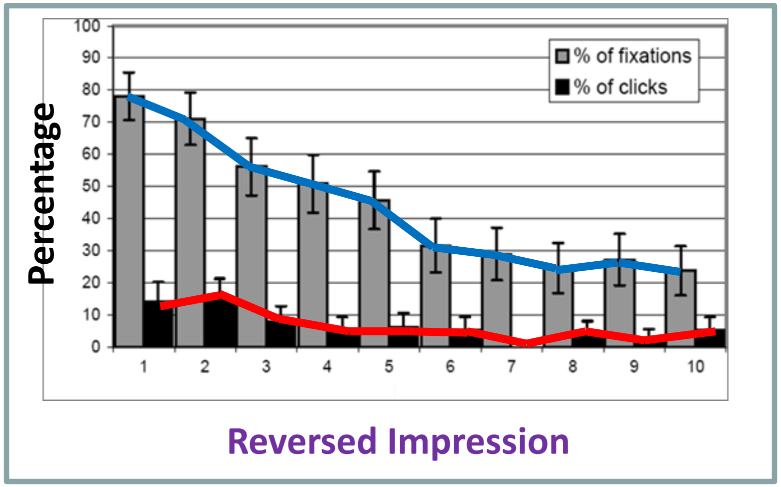
Google user patterns



Click Position-bias



Normal Position



Reversed Impression

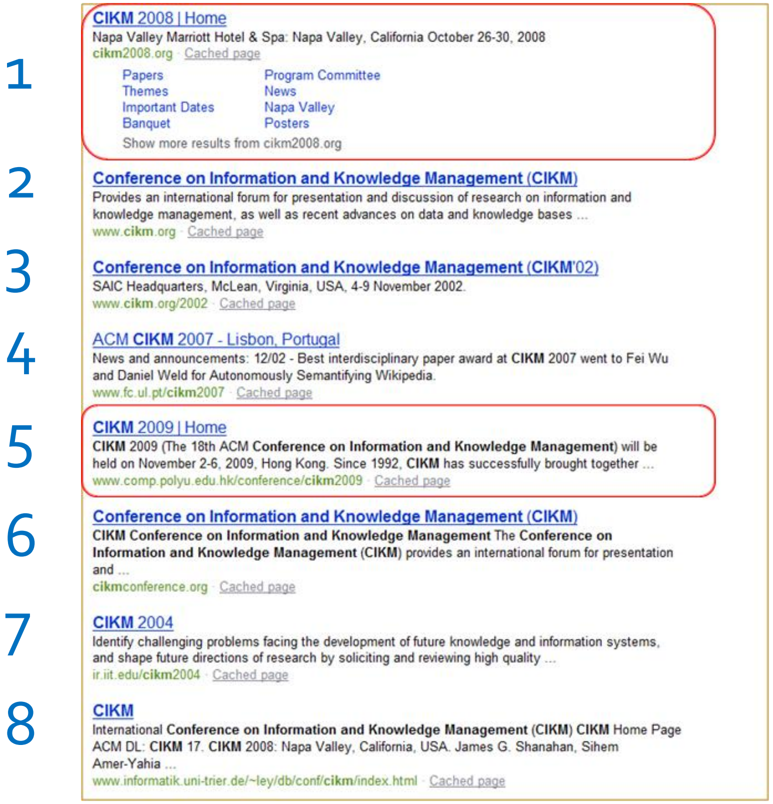
- Higher positions receive more user attention (eye fixation) and clicks than lower positions.
- This is true even in the extreme setting where the order of positions is reversed.
- “Clicks are informative but biased”.

[Joachims+07]

2/23/2015

Clicks as Relative Judgments for Rank Training

- **“Clicked > Skipped Above” [Joachims, KDD02]**



1

CIKM 2008 | Home
Napa Valley Marriott Hotel & Spa: Napa Valley, California October 26-30, 2008
cikm2008.org · Cached page

Papers	Program Committee
Themes	News
Important Dates	Napa Valley
Banquet	Posters

Show more results from cikm2008.org

2

Conference on Information and Knowledge Management (CIKM)
Provides an international forum for presentation and discussion of research on information and knowledge management, as well as recent advances on data and knowledge bases ...
www.cikm.org · Cached page

3

Conference on Information and Knowledge Management (CIKM'02)
SAIC Headquarters, McLean, Virginia, USA, 4-9 November 2002.
www.cikm.org/2002 · Cached page

4

ACM **CIKM** 2007 - Lisbon, Portugal
News and announcements: 12/02 - Best interdisciplinary paper award at **CIKM** 2007 went to Fei Wu and Daniel Weld for Autonomously Semantifying Wikipedia.
www.fc.ul.pt/cikm2007 · Cached page

5

CIKM 2009 | Home
CIKM 2009 (The 18th ACM **Conference on Information and Knowledge Management**) will be held on November 2-6, 2009, Hong Kong. Since 1992, **CIKM** has successfully brought together ...
www.comp.polyu.edu.hk/conference/cikm2009 · Cached page

6

Conference on Information and Knowledge Management (CIKM)
CIKM Conference on Information and Knowledge Management The **Conference on Information and Knowledge Management** (CIKM) provides an international forum for presentation and ...
cikmconference.org · Cached page

7

CIKM 2004
Identify challenging problems facing the development of future knowledge and information systems, and shape future directions of research by soliciting and reviewing high quality ...
ir.iit.edu/cikm2004 · Cached page

8

CIKM
International **Conference on Information and Knowledge Management** (CIKM) **CIKM** Home Page ACM DL: **CIKM** 17. **CIKM** 2008: Napa Valley, California, USA. James G. Shanahan, Sihem Amer-Yahia ...
www.informatik.uni-trier.de/~ley/db/conf/cikm/index.html · Cached page

- Preference pairs: #5>#2, #5>#3, #5>#4.
- Use Rank SVM to optimize the retrieval function.
- Limitation:
  - Confidence of judgments
  - Little implication to user modeling

2/23/2015

Additional relation for relative relevance judgments

click > skip above

last click > click above

click > click earlier

last click > click previous

click > no-click next

Web Search Ranking by Incorporating User Behavior Information Rank pages relevant for a query

- **Eugene Agichtein, Eric Brill, Susan Dumais SIGIR 2006**
- **Categories of Features (Signals) for Web Search Ranking**
  - Content match
    - e.g., page terms, anchor text, term weights, term span
  - Document quality
    - e.g., web topology, spam features
- **Add one more category:**
  - Implicit user feedback from click data

Rich User Behavior Feature Space

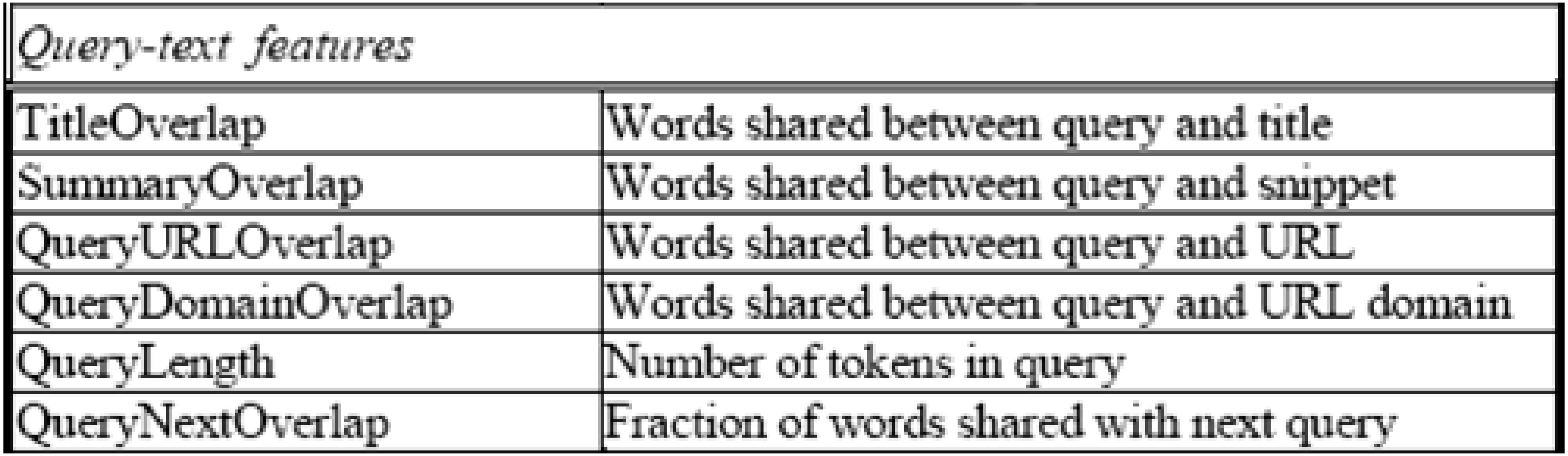
- **Observed and distributional features**
  - Aggregate observed values over all user interactions for each query and result pair
  - Distributional features: deviations from the “expected” behavior for the query

- **Represent user interactions as vectors in user behavior space**
  - **Presentation**: what a user sees *before* a click
  - **Clickthrough**: frequency and timing of clicks
  - **Browsing**: what users do *after* a click

Ranking Features (Signals)

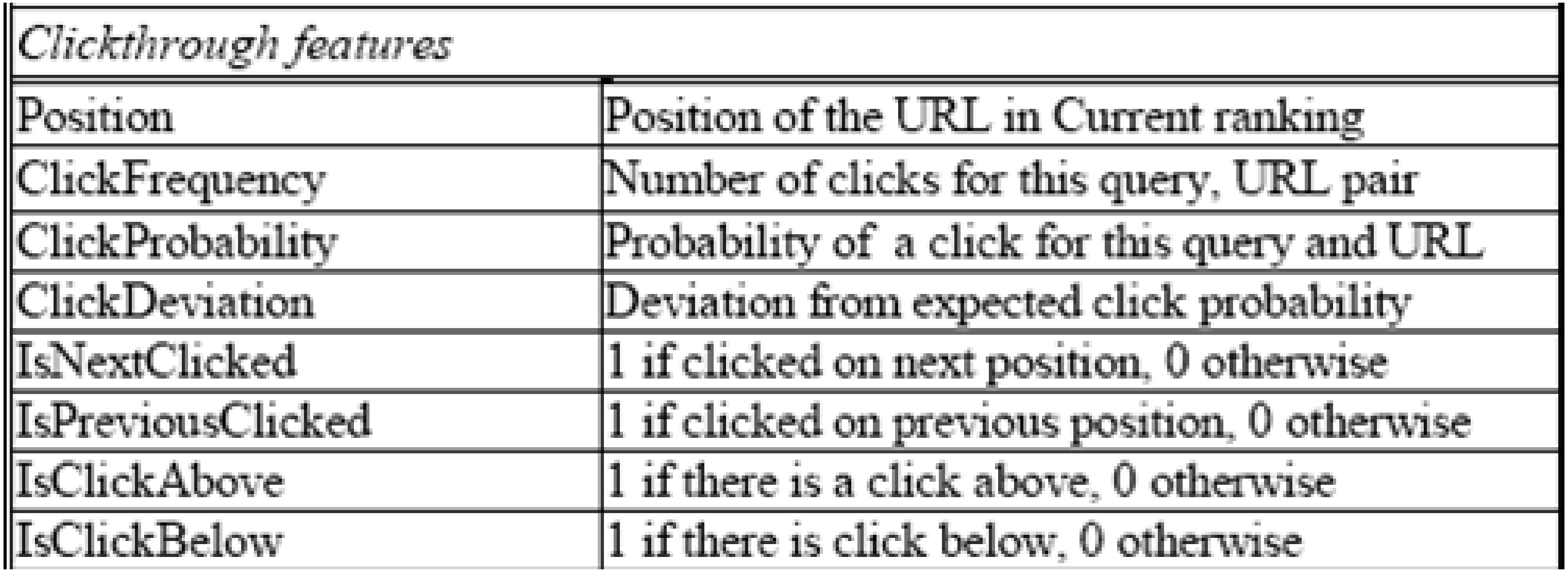
Presentation	
ResultPosition	Position of the URL in Current ranking
QueryTitleOverlap	Fraction of query terms in result Title
Clickthrough	
DeliberationTime	Seconds between query and first click
ClickFrequency	Fraction of all clicks landing on page
ClickDeviation	Deviation from expected click frequency
Browsing	
DwellTime	Result page dwell time
DwellTimeDeviation	Deviation from expected dwell time for query

More Presentation Features



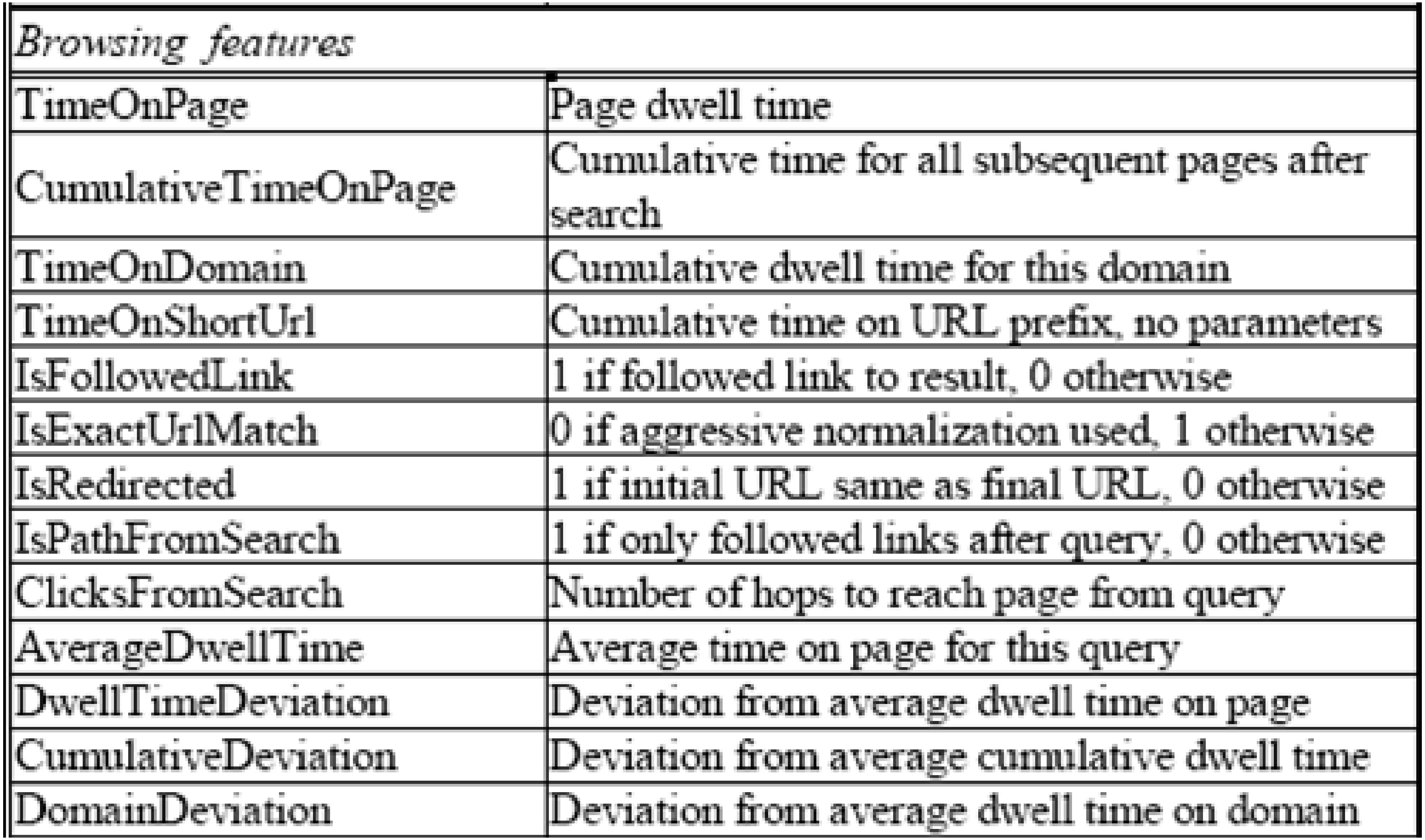
Query-text features	
TitleOverlap	Words shared between query and title
SummaryOverlap	Words shared between query and snippet
QueryURLOverlap	Words shared between query and URL
QueryDomainOverlap	Words shared between query and URL domain
QueryLength	Number of tokens in query
QueryNextOverlap	Fraction of words shared with next query

More Clickthough Features



Clickthrough features	
Position	Position of the URL in Current ranking
ClickFrequency	Number of clicks for this query, URL pair
ClickProbability	Probability of a click for this query and URL
ClickDeviation	Deviation from expected click probability
IsNextClicked	1 if clicked on next position, 0 otherwise
IsPreviousClicked	1 if clicked on previous position, 0 otherwise
IsClickAbove	1 if there is a click above, 0 otherwise
IsClickBelow	1 if there is click below, 0 otherwise

Browsing features



Browsing features	
TimeOnPage	Page dwell time
CumulativeTimeOnPage	Cumulative time for all subsequent pages after search
TimeOnDomain	Cumulative dwell time for this domain
TimeOnShortUrl	Cumulative time on URL prefix, no parameters
IsFollowedLink	1 if followed link to result, 0 otherwise
IsExactUrlMatch	0 if aggressive normalization used, 1 otherwise
IsRedirected	1 if initial URL same as final URL, 0 otherwise
IsPathFromSearch	1 if only followed links after query, 0 otherwise
ClicksFromSearch	Number of hops to reach page from query
AverageDwellTime	Average time on page for this query
DwellTimeDeviation	Deviation from average dwell time on page
CumulativeDeviation	Deviation from average cumulative dwell time
DomainDeviation	Deviation from average dwell time on domain

User Behavior Models for Ranking

- **Use interactions from previous instances of query**
  - General-purpose (not personalized)
  - Only available for queries with past user interactions

- **3 Models:**
  - Rerank results by number of clicks (clickthrough rate)
  - Rerank with all user behavior features).
  - Integrate directly into ranker:
    incorporate user behavior features with other categories of ranking (e.g. text matching)

Evaluation Metrics

- **Precision at K: fraction of relevant in top K**

- **NDCG at K: norm. discounted cumulative gain**
  - Top-ranked results most important

$$N_q = M_q \sum_{j=1}^{K} (2^{r(j)} - 1) / \log(1 + j)$$

- **MAP: mean average precision**
  - Average precision for each query: mean of the precision at K values computed after each relevant document was retrieved

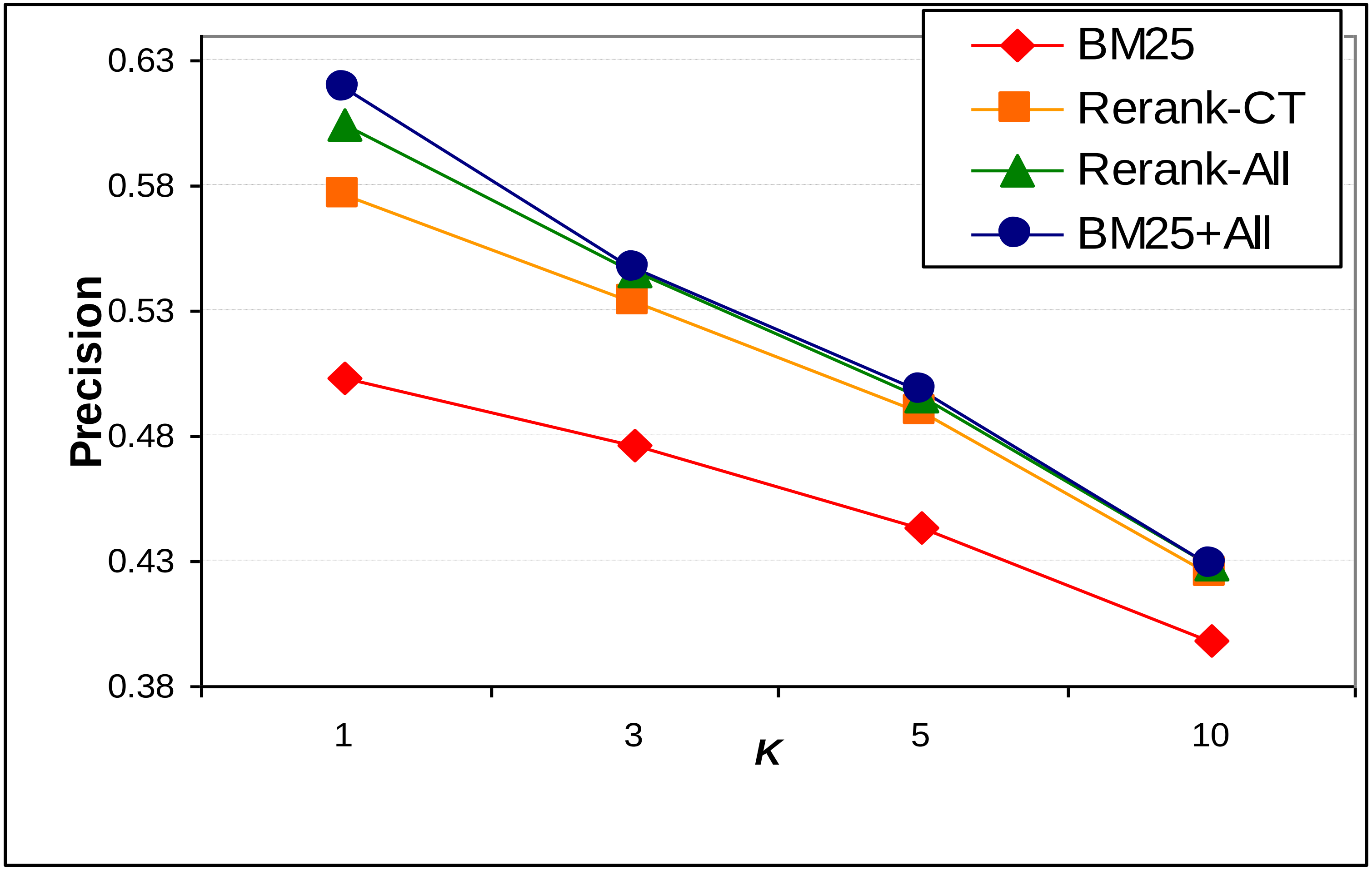
Datasets

- **8 weeks of user behavior data from anonymized opt-in client instrumentation**
- **Millions of unique queries and interaction traces**
- **Random sample of 3,000 queries**
  - Gathered independently of user behavior
  - 1,500 train, 500 validation, 1,000 test
- **Explicit relevance assessments for top 10 results for each query in sample**

Methods Compared

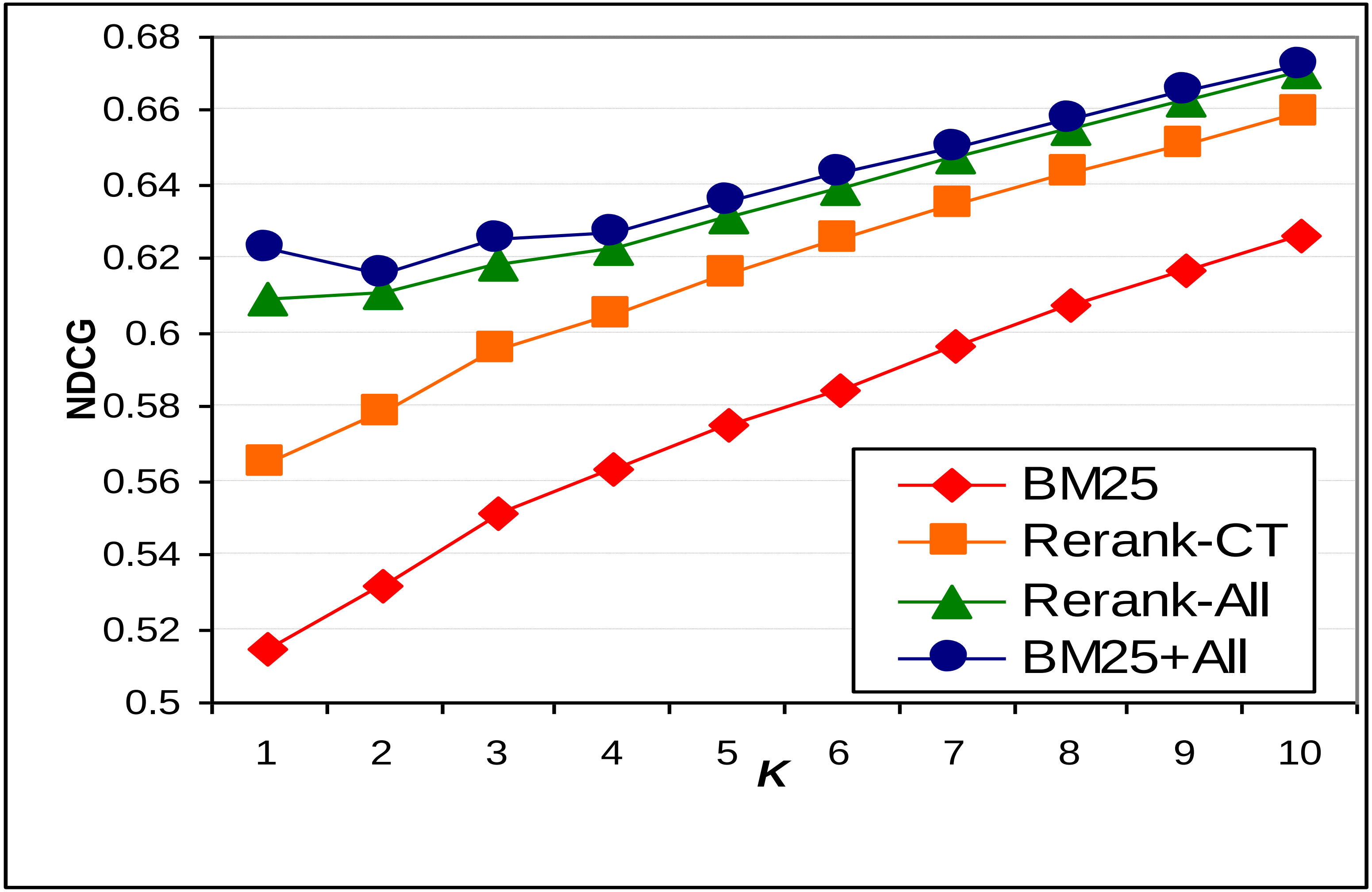
- **Full Search Engine**
  - Content match feature uses BM25F
  - A variation of TF-IDF model
- **Compare 4 ranking models**
  - **BM25F only**
  - Clickthrough: called **Rerank-CT**
    - Rerank these queries with sufficient historic click data
  - Full user behavior model predictions: called **Rerank-All**
  - Integrate all user behavior features directly: **+All**
    - **User behavior features + content match**

Content, User Behavior: Precision at K, queries with interactions



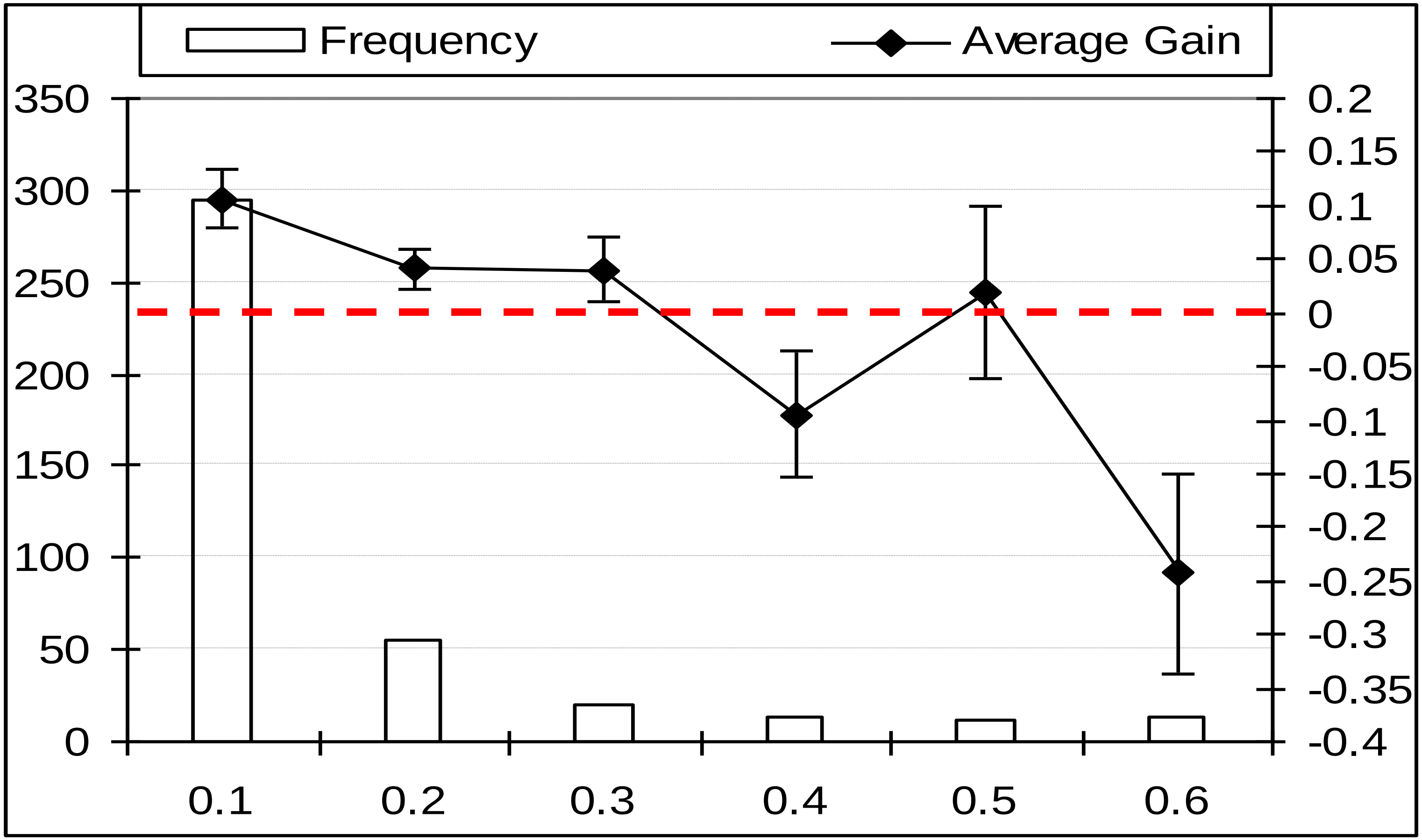
BM25 < Rerank-CT < Rerank-All < +All

Content, User Behavior: NDCG



BM25 < Rerank-CT < Rerank-All < +All

Which Queries Benefit Most



Most gains are for queries with poor ranking

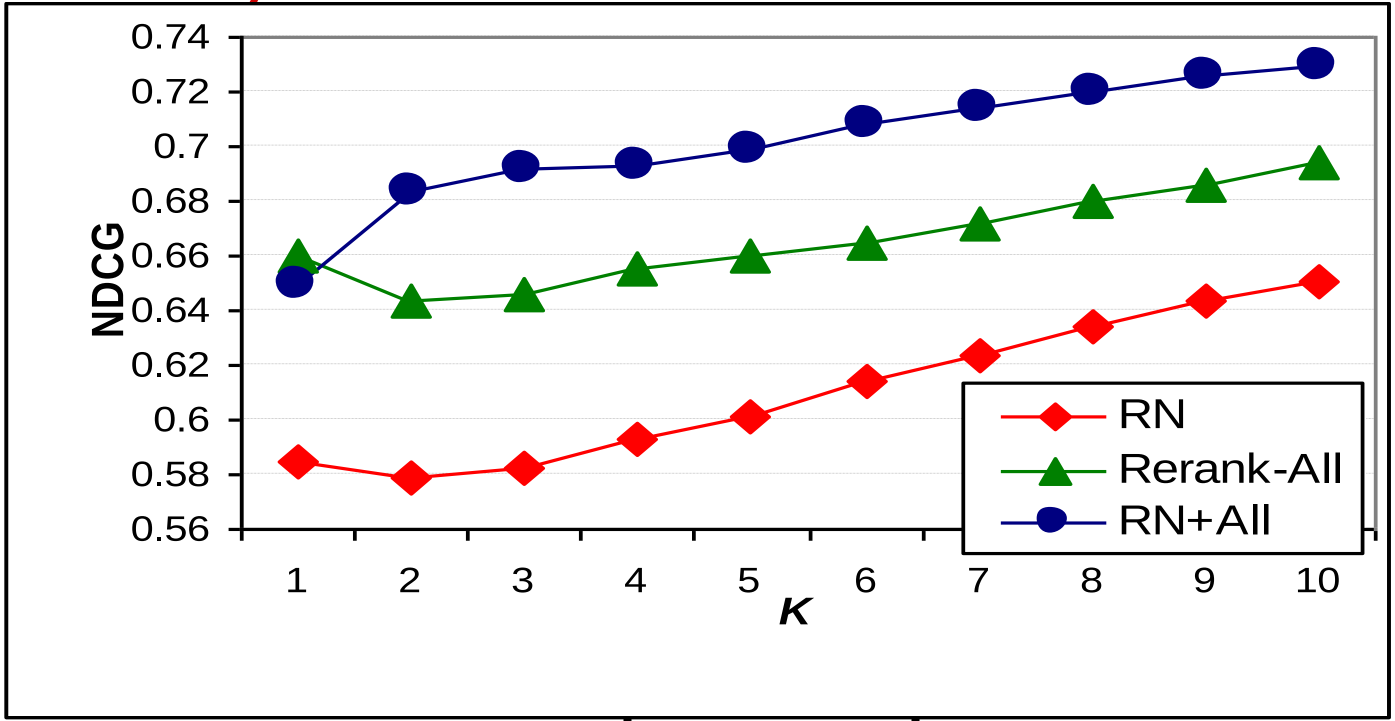
Conclusions

- Incorporating user behavior into web search ranking dramatically improves relevance

- Providing rich user interaction features to ranker is the most effective strategy

- Large improvement shown for up to 50% of test queries

Full Search Engine, User Behavior: NDCG, MAP



RN	0.270	
RN+ALL	0.321	0.052 (19.13%)
BM25	0.236	
BM25+ALL	0.292	0.056 (23.71%)