

**CLEF 2014** Conference and Labs of the Evaluation Forum  
*Information Access Evaluation meets Multilinguality, Multimodality, and Interaction*  
*15 - 18 September 2014, Sheffield - UK*



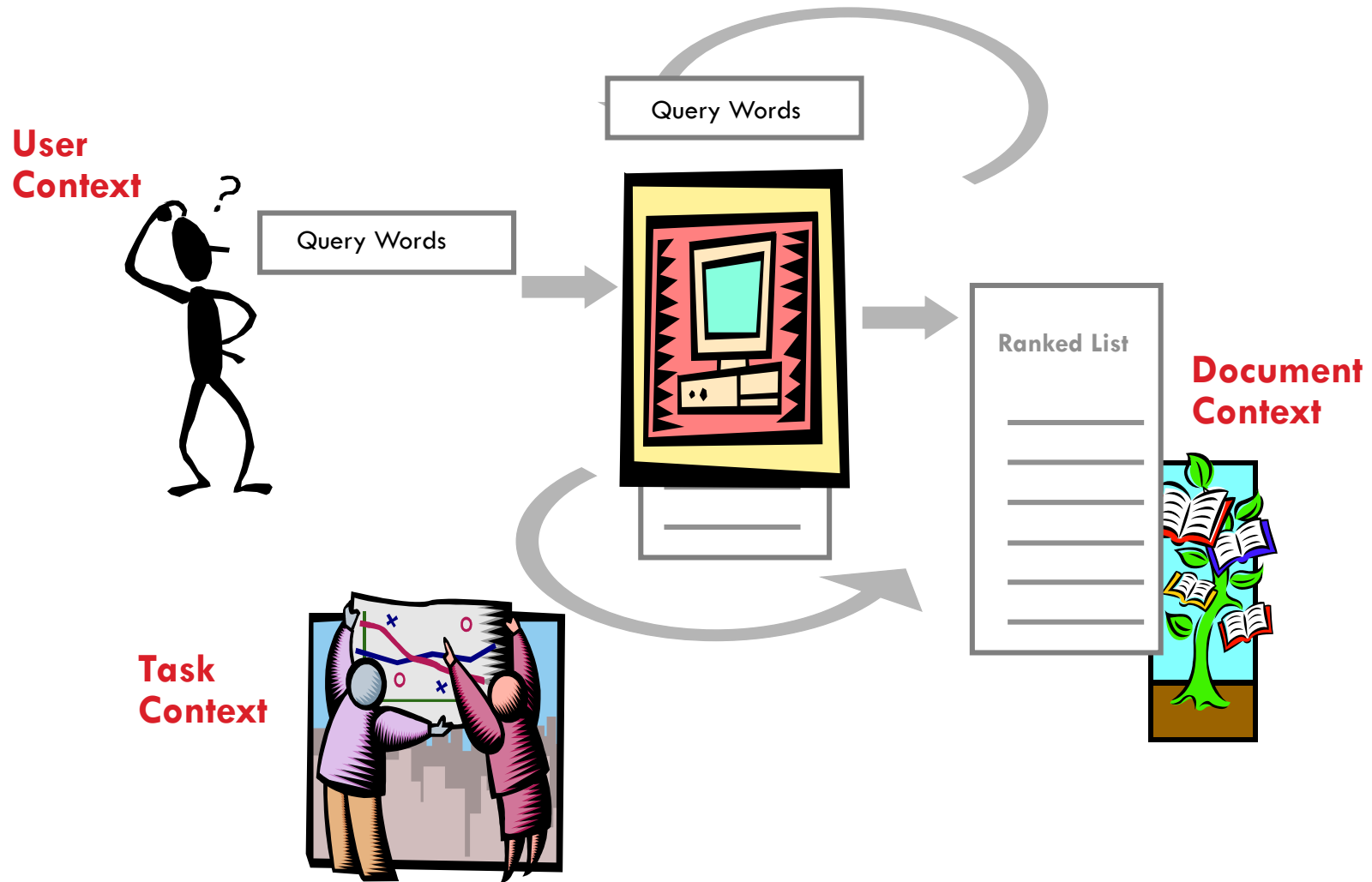
# SEARCH AND CONTEXT

Susan Dumais, Microsoft Research

# Overview

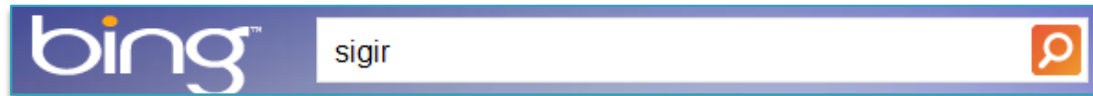
- Importance of context in information retrieval
- “Potential for personalization” framework
- Examples with varied user models and evaluation methods
  - ▣ Personal navigation
  - ▣ Client-side personalization
  - ▣ Short- and long-term models
  - ▣ Time-aware models
- Challenges and new directions

# Search and Context



# Context Improves Query Understanding

- Queries are difficult to interpret in isolation



□ Easier if we can model: who is asking, what they have done in the past, where they are, when it is, etc.

**Searcher:** (*SIGIR* | Susan Dumais ... an information retrieval researcher)

vs. (*SIGIR* | Stuart Bowen Jr. ... the Special Inspector General for Iraq Reconstruction)

**Previous actions:** (*SIGIR* | information retrieval)

vs. (*SIGIR* | U.S. coalitional provisional authority)

**Location:** (*SIGIR* | at SIGIR conference) vs. (*SIGIR* | in Washington DC)

**Time:** (*SIGIR* | Jan. submission) vs. (*SIGIR* | Aug. conference)

- Using a single ranking for everyone, in every context, at every point in time, limits how well a search engine can do

# CLEF 2014

- Have you searched for **CLEF 2014** recently?
- What were you looking for?

[CLEF 2014 | Conference and Labs of the Evaluation Forum](#)

[clef2014.clef-initiative.eu](http://clef2014.clef-initiative.eu) ▼

Welcome. CLEF 2014 is the fifth  
campaigns which have run since

[CLEF 2014 | Conference and Labs of the Evaluation Forum](#)

[clef2014.clef-initiative.eu/index.php?page=Pages/labs.html](http://clef2014.clef-initiative.eu/index.php?page=Pages/labs.html) ▼

CLEF 2014 Conference and Labs of the Evaluation Forum Information Access Evaluation  
Quality, and Interaction 15 - 18 September 2014 ...

[ImageCLEF 2014 | ImageCLEF - Image Retrieval in CLEF](#)

[www.imageclef.org/2014](http://www.imageclef.org/2014) ▼

ImageCLEF - Image Retrieval in CLEF. Navigation. Image  
Image annotation; Liver CT annotation; Domain adaptation

[LifeCLEF 2014 | ImageCLEF - Image Retrieval in CLEF](#)

[www.imageclef.org/lifeclef/2014](http://www.imageclef.org/lifeclef/2014) ▼

ImageCLEF - Image Retrieval in CLEF. Navigation. ImageCLEF 2014; LifeCLEF 2014. ...  
(e tab) Revisions ... ..

[CLEF 2014](#)

[clef2014.clef-initiative.eu/resources](http://clef2014.clef-initiative.eu/resources) ▼

Call for Labs Participation CLEF  
campaign and workshop series

[CLEF 2014 | Conference and Labs of the Evaluation Forum](#)

[clef2014.clef-initiative.eu/index.php?page=Pages/programme.html](http://clef2014.clef-initiative.eu/index.php?page=Pages/programme.html) ▼

CLEF 2014 Conference and Labs of the Evaluation Forum Information Access Evaluation  
Quality, and Interaction 15 - 18 September 2014 ...

[PDF] [Clef Notes 2014](#)

[www.esm.rochester.edu/studentlife/files/Clef-Notes.pdf](http://www.esm.rochester.edu/studentlife/files/Clef-Notes.pdf)

Clef Notes 2014 Eastman School of Music Summer E-Newsletter  
Student Living Center 103, 100 Gibbs Street, Rochester, NY 146

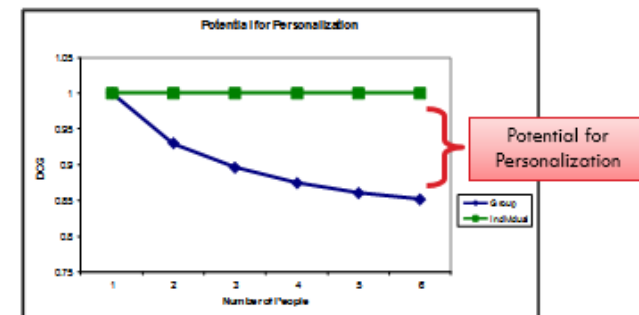
[UNC Clef Hangers | Carolina's Oldest A Capella Group](#)

[clefhangers.com](http://clefhangers.com) ▼

Spring Concert. Memorial Hall UNC-Chapel Hill Saturday, October 25, 2014 \_

# Potential For Personalization

- A single ranking for everyone limits search quality
- Quantify the variation in individual relevance for the same query
- Different ways to measure individual relevance
  - ▣ Explicit judgments from different people for the same query
  - ▣ Implicit judgments (search result clicks, content analysis)
- Personalization can lead to large improvements
  - ▣ Study with explicit judgments
  - ▣ 46% improvements for core ranking
  - ▣ 70% improvements with personalization

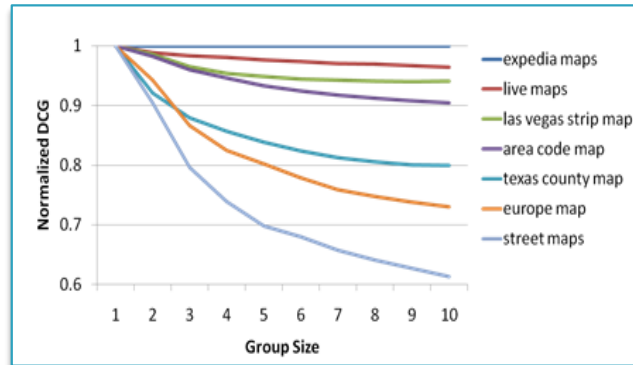


# Potential For Personalization

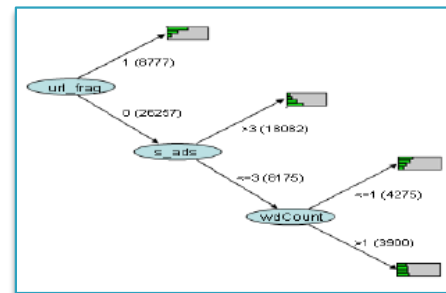
□ Not all queries have high potential for personalization

□ E.g., facebook vs. sigir

□ E.g., \* maps



□ Learn when to personalize



# User Models

- Constructing user models
  - ▣ Sources of evidence
    - Content: Queries, content of web pages, desktop index, etc.
    - Behavior: Visited web pages, explicit feedback, implicit feedback
    - Context: Location, time (of day/week/year), device, etc.
  - ▣ Time frames: Short-term, long-term
  - ▣ Who: Individual, group
- Using user models
  - ▣ Where resides: Client, server
  - ▣ How used: Ranking, query support, presentation, etc.
  - ▣ When used: Always, sometimes, context learned



# User Models

## □ Constructing user models

### ▣ Sources of evidence

- Content: Queries, content of web pages, desktop index, etc.
- Behavior: Visited web pages, explicit feedback, implicit feedback
- Context: Location, time (of day/week/year), device, etc.

### ▣ Time frames: Short-term, long-term

### ▣ Who: Individual, group

PNav

## □ Using user models

### ▣ Where resides: Client, server

### ▣ How used: Ranking, query support, presentation, etc.

### ▣ When used: Always, sometimes, context learned

PSearch

Short/Long

Time

# Example 1: Personal Navigation

- Re-finding is common in Web search
  - ▣ 33% of queries are repeat queries
  - ▣ 39% of clicks are repeat clicks
- Many of these are navigational queries
  - ▣ E.g., facebook -> [www.facebook.com](http://www.facebook.com)
  - ▣ Consistent intent across individuals
  - ▣ Identified via low click entropy
- “Personal navigational” queries
  - ▣ Different intents across individuals, ... but consistently the same intent for an individual
    - SIGIR (for Dumais) -> [www.sigir.org/sigir2014](http://www.sigir.org/sigir2014)
    - SIGIR (for Bowen Jr.) -> [www.sigir.mil](http://www.sigir.mil)

		Repeat Click	New Click
Repeat Query	33%	29%	4%
New Query	67%	10%	57%
		39%	61%

bing  
sigir

446,000 RESULTS

SIGIR Conference is on Sunday, Aug. tomorrow.

ACM SIGIR Special Interest Group on Information Retrieval ...  
www.sigir.org -  
Welcome to the ACM SIGIR Web site. ACM SIGIR addresses issues ranging from theoretical to user demands in the application of computers to the acquisition, organization ...

Welcome to SIGIR | Home  
www.sigir.mil -  
An Iraq fisherman pushes his boat off-shore to depart on his daily fishing trip. View the Report.

home [ACM SIGIR 2010]  
www.sigir2010.org -  
ACM-SIGIR 2010 was held at UniMail, Geneva, Switzerland between 19th and 23rd of July 2010. Thanks to all the participants!! The story continues with ACM-SIGIR 2011.

SIGIR Portland Oregon 2012 - ACM SIGIR Special Interest Group  
www.sigir.org/sigir2012 -  
SIGIR 2012. Online registration for SIGIR 2012 is now closed. On-site registration will be available at the conference venue. Welcome to SIGIR 2012, the SIGIR 2012 ...

Welcome to The 34th Annual ACM SIGIR  
sigir2011.org -  
ACM-SIGIR 2011 successfully completed in Beijing H... participants. Thanks to all the speakers and participants!! See ...

Related searches for sigir  
SIGIR Iraq SIGIR Forum  
SIGIR 12 SIGIR 2011 Accepted  
CIKM WSDM

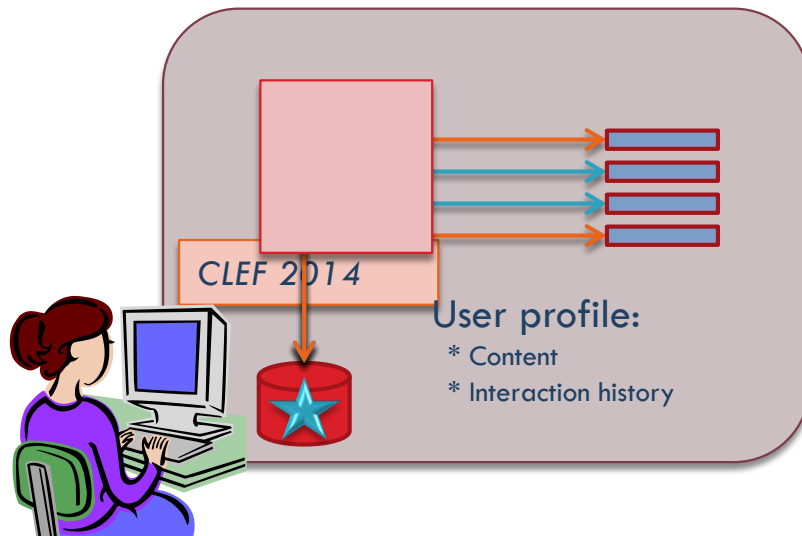
Special Inspector General for Iraq Reconstruction - Wikipedia  
en.wikipedia.org/wiki/Special\_Inspector\_General\_for\_Iraq...  
The Office of the Special Inspector General for Iraq Reconstruction (SIGIR) was created in October 2004 as the successor to the Coalition Provisional Authority Office ...

# Personal Navigation Details

- Large-scale log analysis & online A/B evaluation
- Identifying personal navigation queries
  - ▣ Use consistency of clicks within an individual
  - ▣ Specifically, the last two times a person issued the query, did they have a unique click on same result?
- Coverage and prediction
  - ▣ Many such queries: ~12% of queries
  - ▣ Prediction accuracy high: ~95% accuracy
    - Consistent over time
  - ▣ High coverage, low risk personalization
- Used to re-rank results, and augment presentation

# Example 2: PSearch

- Rich client-side model of a user's interests
  - ▣ Model: Content from desktop search index & Interaction history  
Rich and constantly evolving user model
  - ▣ Client-side re-ranking of (lots of) web search results using model
  - ▣ Good privacy (only the query is sent to server)
    - But, limited portability, and use of community



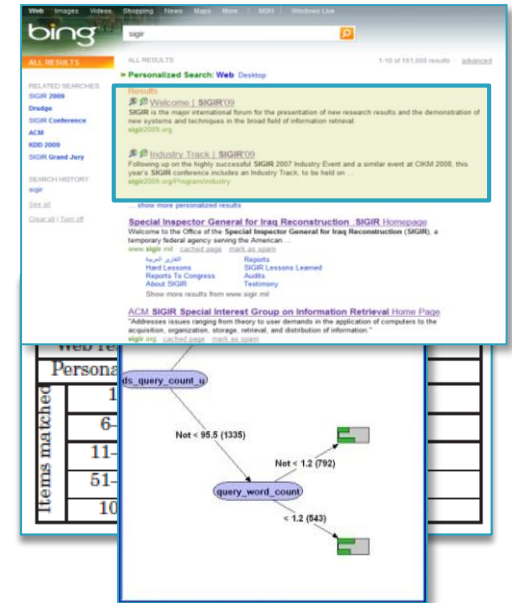
# PSearch Details

## □ Personalized ranking model

- Score: Weighted combination of personal and global web features
  - $Score(result_i) = \alpha PersonalScore(result_i) + (1 - \alpha) WebScore(result_i)$
- Personal score: Content and interaction history features
  - Content score: log odds of term in personal vs. web content
  - Interaction history score: visits to the specific URL, and back off to site

## □ Evaluation

- Offline evaluation, using explicit judgments
- *In situ* evaluation, using PSearch prototype
  - 225+ people for several months
  - Effectiveness:
    - CTR 28% higher, for personalized results
    - CTR 74% higher, when personal evidence is strong
  - Learned model for when to personalize



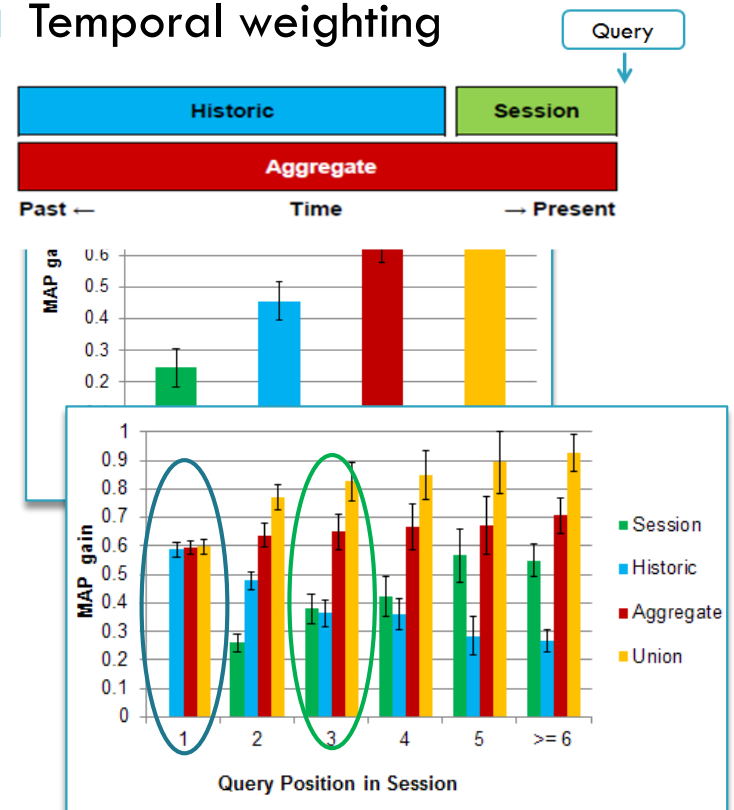
# Example 3: Short + Long

- Short-term context
  - ▣ Previous actions (queries, clicks) within current session
    - (Q=*sigir* | *information retrieval* vs. *iraq reconstruction*)
    - (Q=*ego* | *id* vs. *dangerously in love* vs. *eldorado gold corporation*)
    - (Q=*acl* | *computational linguistics* vs. *knee injury* vs. *country music*)
- Long-term preferences and interests
  - ▣ Behavior: Specific queries/URLs
    - (Q=*weather*) -> *weather.com* vs. *weather.gov* vs. *intellicast.com*
  - ▣ Content: Language models, topic models, etc.
- Learned model to combine both

# Short + Long Details

- User model (content)
  - ▣ Specific queries/URLs
  - ▣ Topic distributions, using ODP
- Which sources are important?
  - ▣ Session (short-term): +25%
  - ▣ Historic (long-term): +45%
  - ▣ Combinations: +65-75%
- What happens within a session?
  - ▣ 60% sessions involve multiple queries
    - 1<sup>st</sup> query, can only use historical
    - By 3<sup>rd</sup> query, short-term features more important than long-term

- User model (temporal extent)
  - ▣ Session, Historical, Combinations
  - ▣ Temporal weighting



# Atypical Sessions

## □ Example user model

55% Football (“nfl”, “philadelphia eagles”, “mark sanchez”)  
 14% Boxing (“espn boxing”, “mickey garcia”, “hbo boxing”)  
 09% Television (“modern family”, “dexter 8”, “tv guide”)  
 06% Travel (“rome hotels”, “tripadvisor seattle”, “rome pasta”)  
 05% Hockey (“elmira pioneers”, “umass lax”, “necbl”)

### New Session 1:

Boxing (“soto vs ortiz k  
 Boxing (“humberto soto”)

**Typical**

### New Session 2:

Dentistry (“oral sores”)  
 Dentistry (“aphthous sore”)  
 Healthcare (“aphthous ulcer treatment”)

**Atypical**

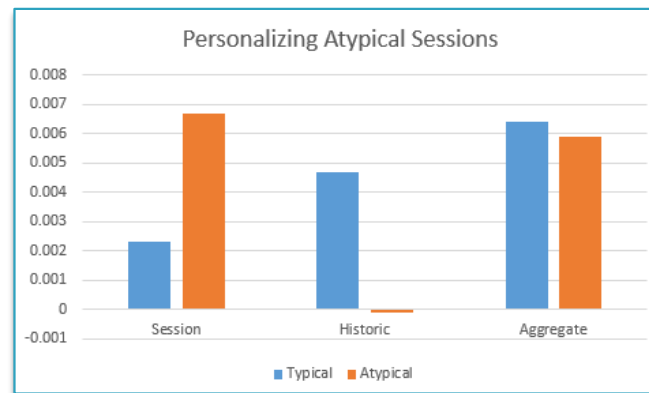
## □ ~6% of session atypical

- ▣ Tend to be more complex, and have poor quality results
- ▣ Common topics: Medical (49%), Computers (24%)
- ▣ What you need to do vs. what you choose to do



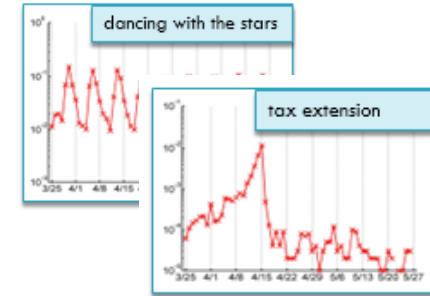
# Atypical Sessions Details

- Learn model to identify atypical sessions
  - ▣ Logistic regressions classifier
- Apply different personalization models for them
  - ▣ If typical, use long-term user model
  - ▣ If atypical, use short-term session user model
- Accuracy by similarity of session to user model



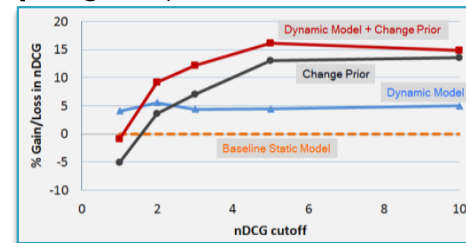
# Example 4: Temporal Dynamics

- Queries are not uniformly distributed over time
  - ▣ Often triggered by events in the world
- What's relevant changes over time
  - ▣ E.g., *US Open* ... in 2014 vs. in 2013
  - ▣ E.g., *US Open 2014* ... in May (golf) vs. in Sept (tennis)
  - ▣ E.g., *US Tennis Open 2014* ...
    - Before event: Schedules and tickets, e.g., stubhub
    - During event: Real-time scores or broadcast, e.g., espn
    - After event: General sites, e.g., wikipedia, usta

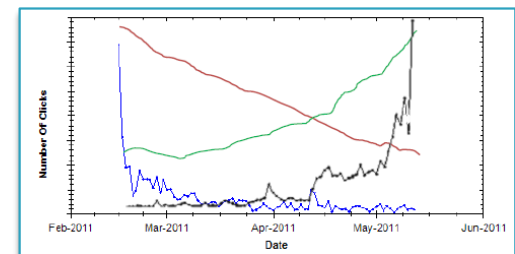


# Temporal Dynamics Details

- Develop time-aware retrieval models
- Model content change on a page
  - ▣ Pages have different *rates of change* (influences document priors,  $P(D)$ )
  - ▣ Terms have different *longevity* on a page (influences term weights,  $P(Q|D)$ )
  - ▣ 15% improvement vs. LM baseline



- Model user interactions as a time-series
  - ▣ Model Query and URL clicks as time-series
  - ▣ Enables appropriate weighting of historical interaction data
  - ▣ Useful for queries with local or global trends



# Challenges in Personalization

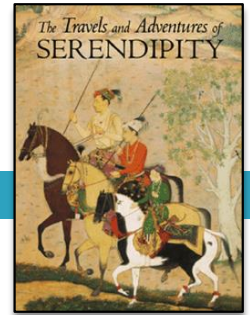
- User-centered
  - ▣ Privacy
  - ▣ Transparency and control
  - ▣ Serendipity
  
- Systems-centered
  - ▣ Evaluation
    - Measurement, experimentation
  - ▣ System optimization
    - Storage, run-time, caching, etc.

# Privacy



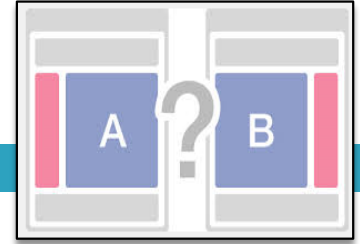
- ❑ User profile and content need to be in the same place
- ❑ Local profile (e.g., PSearch)
  - ❑ Local profile, local computation
  - ❑ Only query sent to server
- ❑ Cloud profile (e.g., Web search)
  - ❑ Cloud profile, cloud computation
  - ❑ Transparency and control over what's stored
- ❑ Other approaches
  - ❑ Light weight profiles (e.g., queries in a session)
  - ❑ Public or semi-public profiles (e.g., tweets, Facebook status)
  - ❑ Matching to a group vs. an individual

# Serendipity



- Does personalization mean the end of serendipity?
  - ▣ ... Actually, it can improve it!
- Experiment on *Relevance vs. Interestingness*
  - ▣ Personalization finds more relevant results
  - ▣ Personalization also finds more interesting results
    - Even when interesting results were not relevant
- Need to be ready for serendipity
  - ▣ ... Like the Princes of Serendip

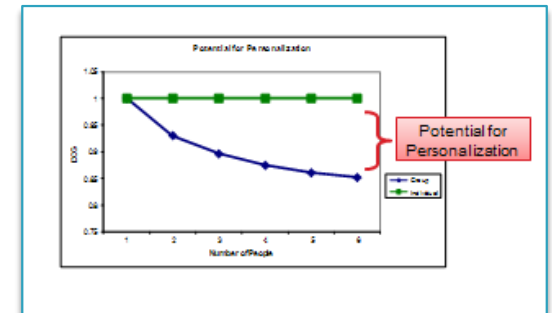
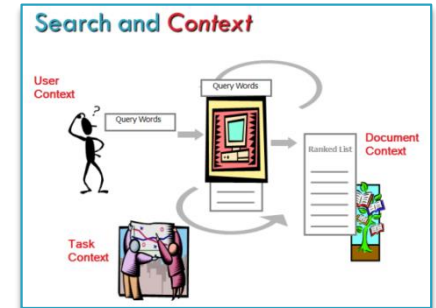
# Evaluation



- External judges, e.g., “assessors”
  - ▣ Lack diversity of intents and realistic context
  - ▣ Crowd workers may help some
- Actual searcher
  - ▣ Offline
    - Allows safe exploration of many different alternatives
    - Labels can be explicit or implicit judgments (log analysis)
  - ▣ Online
    - Explicit judgments: Nice, but annoying and may change behavior
    - Implicit judgments: Scalable, but can be very noisy
    - Note ... not directly repeatable; requires production-level code; mistakes costly; biased toward what is presented; etc.
- Diversity of methods important
  - ▣ User studies, log analysis, and A/B testing

# Summary

- Queries difficult to interpret in isolation
- Augmenting query with context helps
  - ▣ Who, what, where, when?
- Potential for improving search using context is large
- Examples
  - ▣ PNav, PSearch, Short/Long, Time
- Challenges and new directions
  - ▣ Spatio-temporal especially in mobile, social, proactive





# Thanks!

- Questions?

- More info:

<http://research.microsoft.com/~sdumais>

- Collaborators:

- ▣ Eric Horvitz, Jaime Teevan, Paul Bennett, Ryen White, Kevyn Collins-Thompson, Peter Bailey, Eugene Agichtein, Krysta Svore, Kira Radinsky, Jon Elsas, Sarah Tyler, Alex Kotov, Anagha Kulkarni, Paul André, Carsten Eickhoff

# References

## □ Short-term models

- White et al., CIKM 2010. *Predicting short-term interests using activity based contexts.*
- Kotov et al., SIGIR 2011. *Models and analyses of multi-session search tasks.*
- Eickhoff et al., WSDM 2013. *Personalizing atypical search sessions.*
- P. André et al., CHI 2009. [From x-rays to silly putty via Uranus: Serendipity and its role in Web search.](#)

## □ Long-term models

- Teevan et al., SIGIR 2005. *Personalizing search via automated analysis of interests and activities.* \*
- Teevan et al., SIGIR 2008. *To personalize or not: Modeling queries with variations in user intent.* \*
- Teevan et al., TOCHI 2010. *Potential for personalization.* \*
- Teevan et al., WSDM 2011. *Understanding and predicting personal navigation.* \*
- Bennett et al., SIGIR 2012. *Modeling the impact of short- & long-term behavior on search personalization.* \*

## □ Temporal models

- Elsas & Dumais, WSDM 2010. *Leveraging temporal dynamics of document content in relevance ranking.* \*
- Kulkarni et al., WSDM 2011. *Understanding temporal query dynamics.*
- Radinsky et al., TOIS 2013. *Behavioral dynamics on the web: Learning, modeling and predicting.* \*
- [http://www.bing.com/community/site\\_blogs/b/search/archive/2011/02/10/making-search-yours.aspx](http://www.bing.com/community/site_blogs/b/search/archive/2011/02/10/making-search-yours.aspx)
- [http://www.bing.com/community/site\\_blogs/b/search/archive/2011/09/14/adapting-search-to-you.aspx](http://www.bing.com/community/site_blogs/b/search/archive/2011/09/14/adapting-search-to-you.aspx)