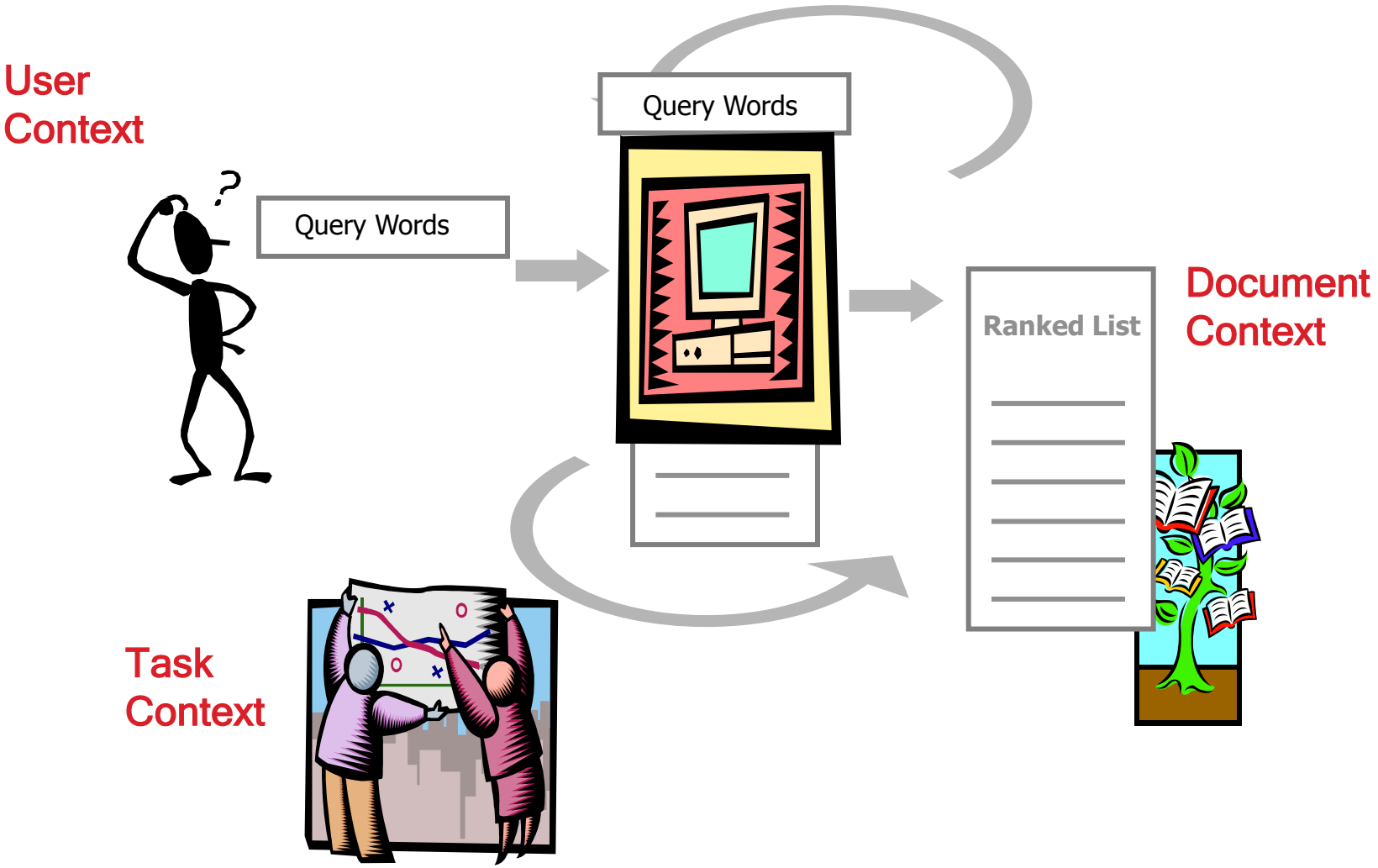
PUTTING CONTEXT INTO SEARCH AND SEARCH INTO CONTEXT

Susan Dumais, Microsoft Research

Overview

- Importance of context in IR
- Potential for personalization framework
- Examples
  - Personal navigation
  - Client-side personalization
  - Short- and long-term models
  - Time as metadata
- 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

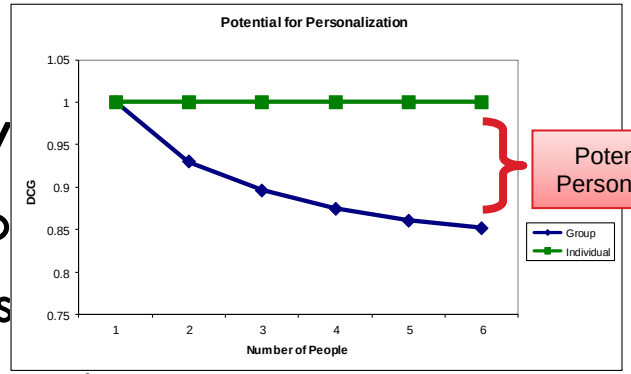
SIGIR 2012?

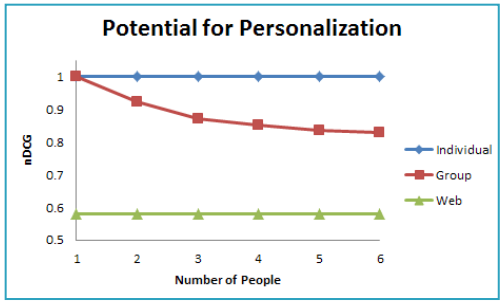
- Have you searched for SIGIR 2012 recently?
- What were you looking for?

Teevan et al., *ToCHI 2010*

Potential For Personalization

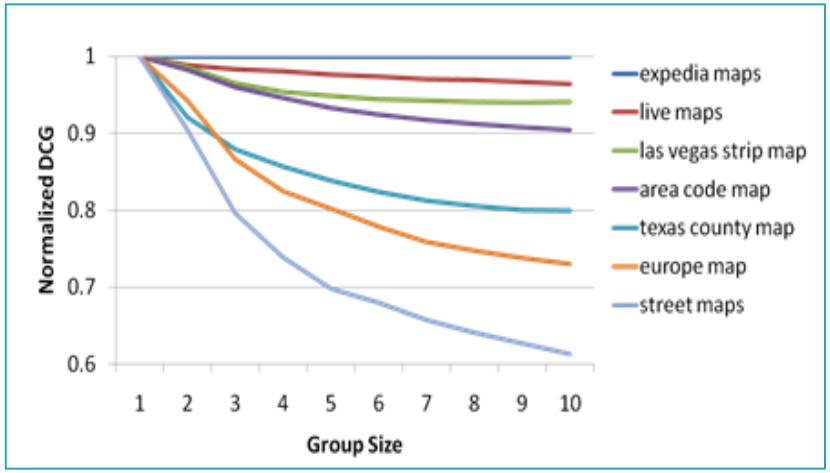
- A single ranking for everyone limits search quality
- Quantify the variation in individual relevance for the same query
- Different ways to measure indiv
  - Explicit judgments from different peo
  - Implicit judgments (search result clicks
- Personalization can lead to large improvements
  - Small study with explicit judgments
  - 46% improvements for core ranking
  - 70% improvements with personalization



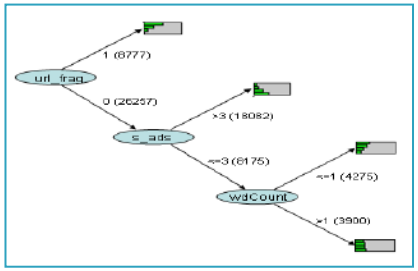


Potential For Personalization

- 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
  - 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
- Using user models
  - Where resides: Client, server
  - How used: Ranking, query support, presentation
  - When used: Always, sometimes, context learned

PNav

PSearch

Short/Long

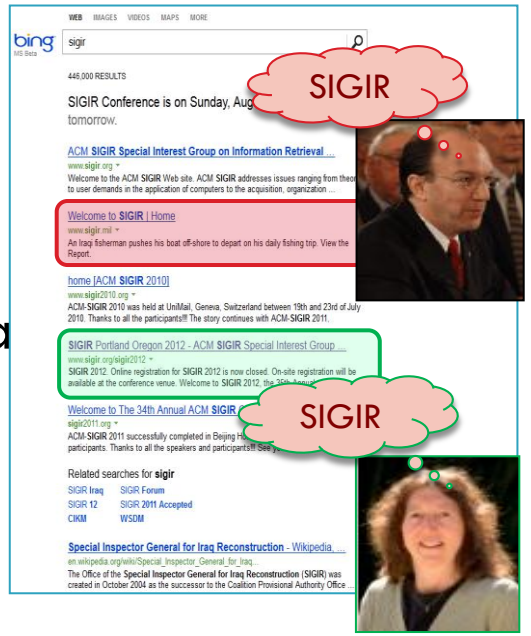
Time

Teevan et al., *SIGIR 2007, WSDM 2010*

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., *microsoft* -> www.microsoft.com
  - Consistent intent across individuals
  - Identified via low click entropy
- "Personal navigational" queries
  - Different intents across individuals, but consistently the same intent for an individua
    - *SIGIR* (for Dumais) -> www.sigir.org/sigir2012
    - *SIGIR* (for Bowen Jr.) -> www.sigir.mil

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

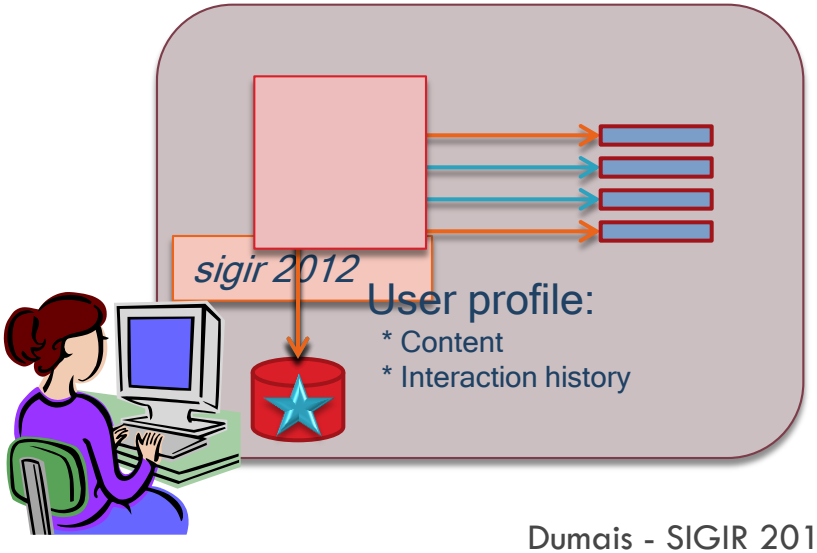


Personal Navigation Details

- Large-scale log analysis
- 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: ~15% of queries
  - Prediction accuracy high: ~95% accuracy
  - High coverage, low risk type of personalization
- Predictions consistent over time
- Can be used to re-rank, or augment presentation

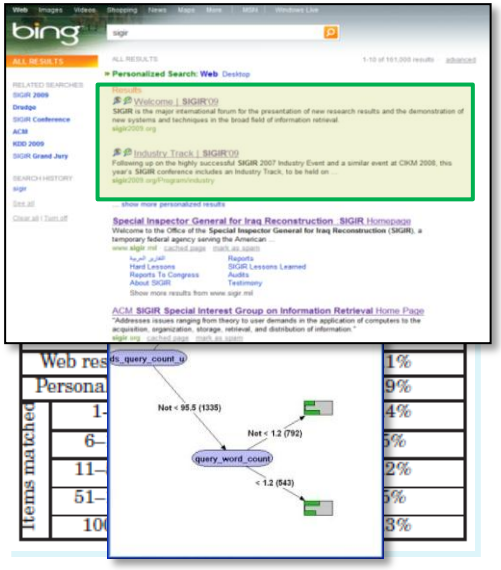
Example 2: PSearch

- 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

- 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 backoff to site
- Evaluation
  - Offline evaluation, using explicit judgments
  - In situ evaluation, using PSearch prototype
    - Internal deployment; 225+ people for several months
    - Coverage: Results personalized for 64% of queries
    - Effectiveness:
      - CTR 28% higher, for personalized results
      - CTR 74% higher, when personal evidence is strong
    - Learned model for when to personalize



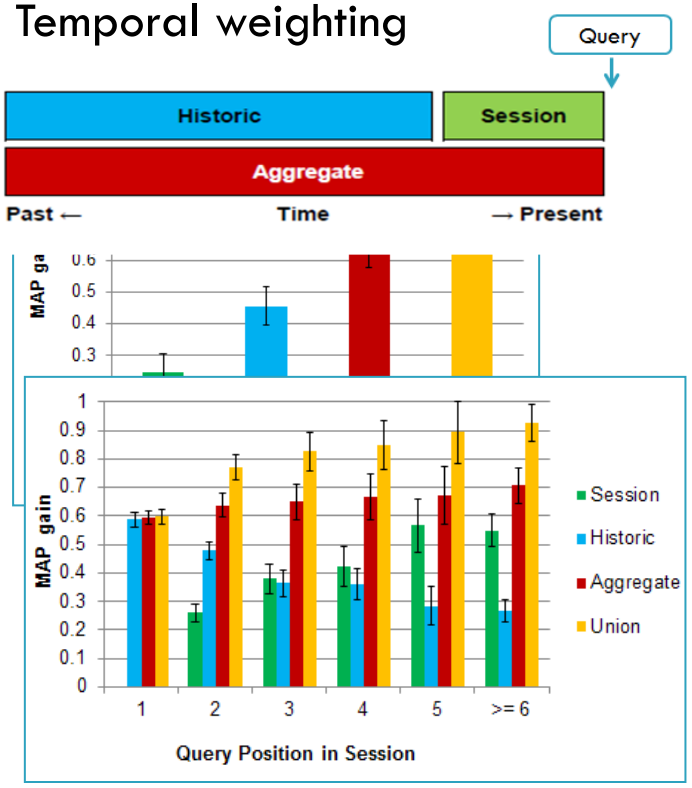
Example 3: Short + Long

- Short-term context
  - Previous actions (queries, clicks) within current session
    - (Q=*sigir* | *information retrieval* vs. *iraq reconstruction*)
    - (Q=*ego* | *id*
    - (Q=*acl* | *computational linguistics*
- 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.
- Develop unified model for both

Short + Long Details

- User model (content)
  - Specific queries/URLs
  - Topic distributions, using ODP
- Log-based evaluation, MAP
- Which sources are important?
  - Session (short-term): +25%
  - Historic (long-term): +45%
  - Combinations: +65-75%
- What happens within a session?
  - 60% of sessions involve multiple queries
    - By 3rd query in session, short-term features more important than long-term
    - First queries in session are different

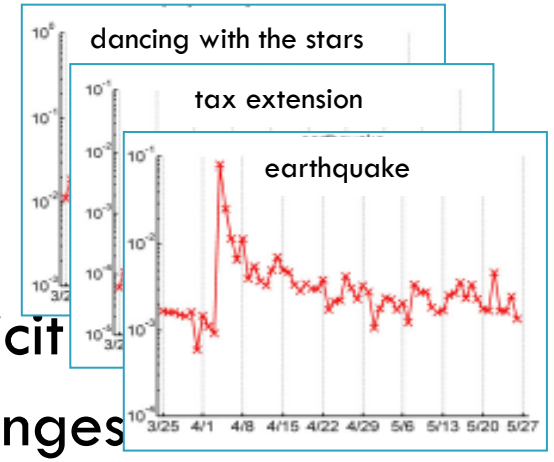
- User model (temporal extent)
  - Session, Historical, Combinations
  - Temporal weighting



Elsas & Dumais, *WSDM 2010*
Radinski et al., *WWW 2012*

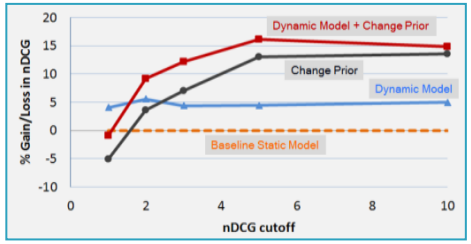
Example 4: Temporal Dynamics

- Queries are not uniformly distributed over time
  - Often triggered by events in the world
- Relevance is influenced by time
  - Explicit time (e.g., *US Open 2012*)
  - Implicit time (e.g., *Olympic results*; implicit
  - What's relevant to the same query changes
    - E.g., *US Open* … in 2012 vs. in 2011
    - E.g., *US Open 2012* … in May (golf) vs. in Sept (tennis)
    - E.g., *US Tennis Open 2012* …
      - Before event: Schedules and tickets, e.g., stubhub
      - During event: Real-time scores or broadcast, e.g., espn, cbssports
      - After event: General sites, e.g., wikipedia, usta

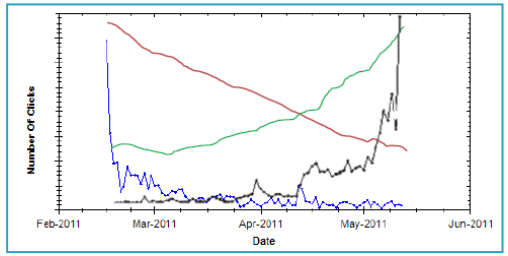


Temporal Dynamics Details

- Develop time-aware retrieval models
- Leverage 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



- Leverage time-series modeling of user interactions
  - 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
  - Consistency
  - Serendipity
- Systems-centered
  - System optimization
    - Storage, run-time, caching, etc.
  - Evaluation

Privacy

- Need user profile and content to be in the same place
- Profile on client (e.g., PSearch)
  - Profile is private
  - Query to server, many documents returned, local computations
- Profile in cloud
  - Transparency about what's stored
  - Control over what's stored … including nothing
- Other possible approaches
  - Light weight profiles (e.g., queries in a session)
  - Public or semi-public profiles (e.g., Tweets, Facebook status)
  - Matching an individual to group

André et al., *CHI 2009*

Serendipity

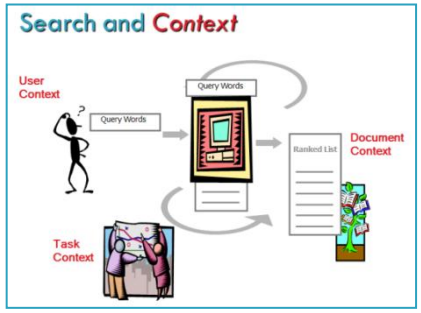
- Does personalization mean the end of serendipity?
- Actually … it can improve it!
- Judgments of Relevance vs. Interestingness
  - Personalization finds more relevant results
  - Personalization finds more interesting results
  - Even when interesting results were not relevant
- Need to be ready for serendipity
  - Zone of proximal learning
  - Walpole's *Three Princes of Serendip* – heroes made discoveries by accident and sagacity, of things they were not in quest of

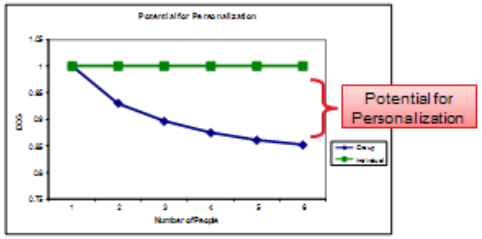
Evaluation

- External judges
  - Query – Lack diversity of intents and backgrounds
  - Query + user profile (e.g., session data) – Better, but where do the profiles come from and how do we summarize them?
- Actual searchers
  - Offline
    - Allows exploration of many different alternatives
    - But … Difficult to collect at scale
  - Online *(In Situ)*
    - Explicit judgments – Great, but annoying and may change behavior
    - Implicit judgments – Nice, but can be noisy
    - But … Limited set of alternatives; presentation and relevance coupled
- Diversity of methods: User studies; user panels; large-scale log analysis and A/B testing

Summary

- Queries difficult to interpret in isolation
- Augmenting query with context can help
  - Who, what, where, when?
- Potential for improving search using context is large
- Examples
- Challenges and new directions





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 Radinski, Jon Elsas, Sarah Tyler, Alex Kotov, Anagha Kulkarni

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.*
  - Agichtein et al., SIGIR 2012. *Search interrupted: Understanding and predicting search task continuation.*
- Long-term models
  - Teevan et al., SIGIR 2005. *Personalizing search via automated analysis of interests and activities.* *
  - Teevan et al. SIGIR 2008. *
  - Teevan et al., TOCHI 2010. *Potential for personalization.* *
  - Bennett et al., SIGIR 2012. *Modeling the impact of short- and long-term behavior on search personalization.* *
  - Tyler et al., WSDM 2010. *Large Scale Query Log Analysis of Re-Finding.*
  - Teevan et al., WSDM 2011. *Understanding and Predicting Personal Navigation.* *
- Time
  - Elsas and Dumais, WSDM 2010. *Leveraging temporal dynamics of document content in relevance ranking.* *
  - Kulkarni et al., WSDM 2011. *Understanding temporal query dynamics.*
  - Radinsky et al., WWW 2012. *Modeling and predicting behavioral dynamics on the Web.* *