Microsoft Research
FacultySummit

Text and Context: Using Context to Better Understand Searchers' Intentions

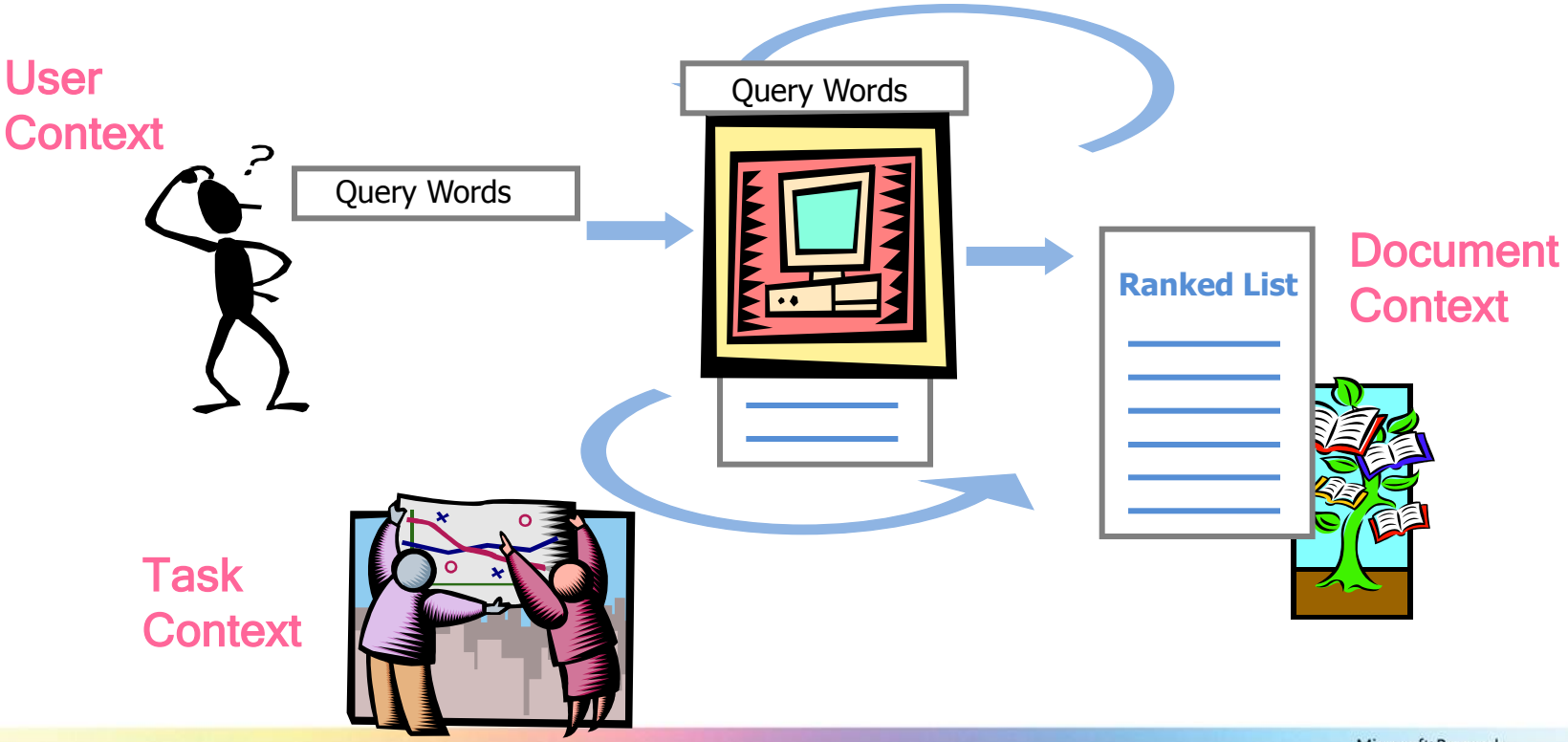
Susan Dumais
Principal Researcher
Microsoft Research

FUTURE/WORLD 2011 2031

Overview

- Understanding a searcher's intention is difficult
  - 20 billion Web pages, given a 2.5 word query!
- Automatic query analysis and reformulation helps
  - Spelling correction, Stemming, Synonym expansion, Phrase identification, Term weighting, etc.
- Augmenting text with context is important
  - Who, what, where, when?
  - Why are you asking?
  - Iterative and evolving “dialog”

Search and Context



Using Context to Improve Query Understanding

- Queries difficult to interpret in isolation
  - E.g., *SIGIR*

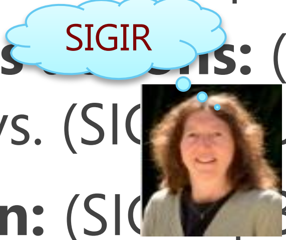
- Easier if we can model: <u>who</u> is asking, <u>what</u> they have done in the past, <u>where</u> they are, <u>when</u> it is, etc.

Searcher: (SIGIR | Susan Dumais) vs. (SIGIR | Stuart Bowen Jr.)

Previous … **s:** (S … natio … eval)

vs. (SI … l provisio …

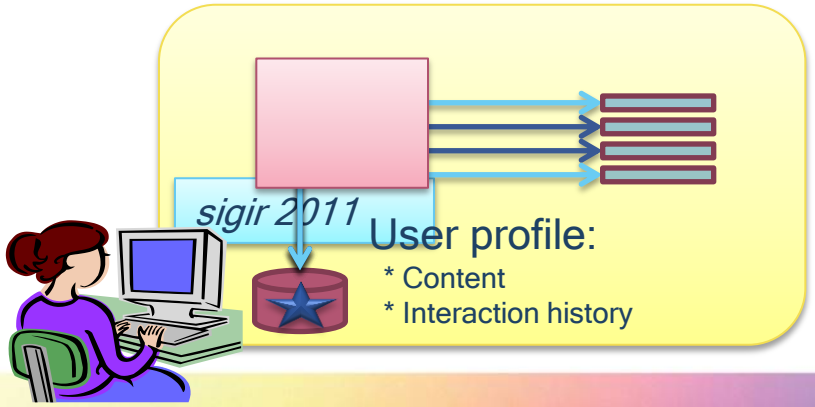
Location: (SI … week) vs. …



Time: (SIGIR | January) vs. (SIGIR | May) vs. (SIGIR | July)

Long-term models (e.g., PSearch)

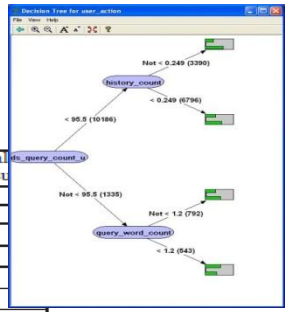
- Single ranking for everyone limits search accuracy
  - "Potential for personalization" framework
- PSearch, client-side model of a user's interests to personalize search
  - Model: Content (desktop search index) and Interaction history
  - Rich and constantly evolving user model
  - Good privacy (only the query is sent to server)
  - But, limited portability, and use of community



PSearch Details

- Ranking Model
  - Personal score: Content and interaction history features
    - Content score based on tf-idf ideas (i.e., log odds of term in personal vs. web content)
    - Interaction history based on visits to the specific URL as well as backoff to site
  - Final score: Weighted combination of personal and global web features
    - $Score(result_i) = \alpha PersonalScore(result_i) + (1 - \alpha) WebScore(result_i)$
- Evaluation
  - Offline evaluation, using explicit judgments
    - Examined alternative corpus, user and document representations
  - In situ evaluation, using PSearch prototype
    - Internal deployment with >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

		Personalized Result Clicks	% of total Queries Issu
Web results		4.3%	36.1%
Personalized		5.5%	63.9%
Items matched	1-5	4.2%	22.4%
	6–10	5.2%	8.5%
	11–50	6.0%	17.2%
	51–100	5.6%	5.5%
	100+	7.5%	10.3%



Short-term models (e.g., session actions)

- Search behavior resides within a short-term context
  - For example, previous actions within the current session
  - This context important for query understanding
    - Query *[sigir]* … given *[information retrieval]* vs. *[iraq reconstruction]*
    - Query *[ego]* … given *[id]* vs. *[dangerously in love]* vs. *[eldorado gold corporation]*
    - Query *[acl]* … given *[computational linguistics]* vs. *[knee injury]* vs. *[country music]*
- Represent queries and URL visits as distributions over ODP classes

- Use for prediction, re-ranking, query suggestion, task support, etc.

Session Details

- Context helps
  - Using any context source improves accuracy
  - Using more sources improves accuracy

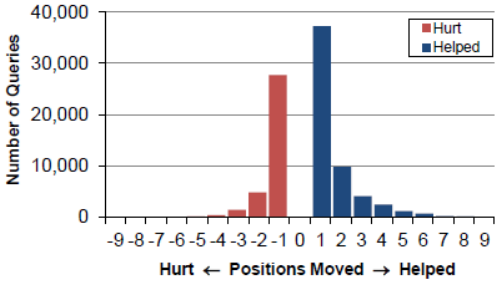
Context source	Accuracy ($F1$)		
	Models		
	Query	Context	Intent
None (i.e., current query only)	0.39	–	0.39
Queries (i.e., all previous queries)	0.39	0.42	0.43
Queries + ***SERPClicks*** (i.e., all previous queries / result clicks)	0.39	0.46	0.46
Queries + ***SERPClicks*** + ***NavTrails*** (i.e., all previous actions)	0.39	0.50	0.49

- Differences across queries
  - Query model wins: current query has specific intent [*espn*], [*webmd*] or first action after a shift in interests
  - Context model wins: query is ambiguous [*amazon*] and session has a consistent intent
  - Intent model wins: session has consistent intent throughout

Context source	Percentage of queries best		
	Between models		
	Query	Context	Intent
Queries (i.e., all previous queries)	25%	18%	22%
Queries + ***SERPClicks*** (i.e., all previous queries / result clicks)	30%	16%	25%
Queries + ***SERPClicks*** + ***NavTrails*** (i.e., all previous actions)	34%	11%	30%

Location

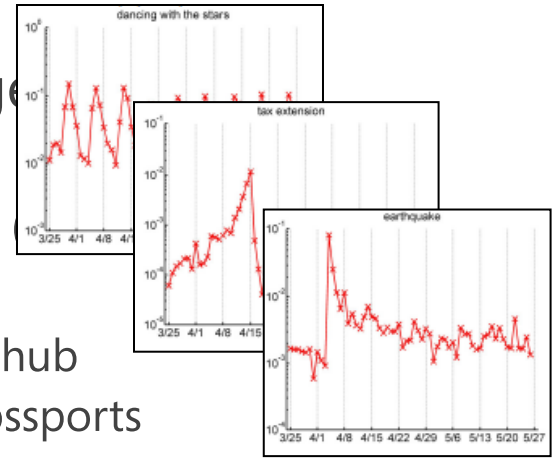
- How much does knowing location help search?
  - Search: H(URL | Query) = 2.8
  - Search & Location: H(URL | Query, IP) = 1.2
- Explicit location (e.g., *susan dumais kirkland wa*)
- Implicit local (e.g., *pizza hut*; implicit "near me")
- Potential for "localization"
  - *SMH*: Sarasota Mem Hospital
  - *LATimes*: local news section



- Mobile searches situated in a location (evolving over time)

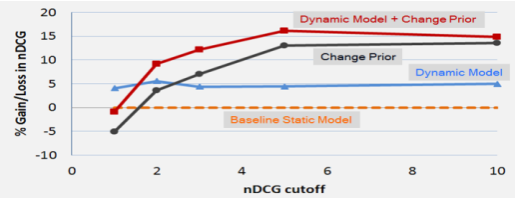
Temporal Dynamics

- Explicit time (e.g., *World Cup Soccer 2011*)
- Implicit time (e.g., *World Cup Soccer*; implicit "now")
- Queries are not uniformly distributed over time
  - Often triggered by events in the world
- What's relevant to the same query change
  - E.g., *Stanley Cup* in 2011 vs. in 2010
  - E.g., *US Open 2011* in May (golf) vs. in Sept
  - E.g., *March madness 2011*
    - Before event: Schedule and tickets, e.g., stubhub
    - During event: Real-time scores, e.g., espn, cbssports
    - After event: General sites, e.g., wikipedia, ncaa



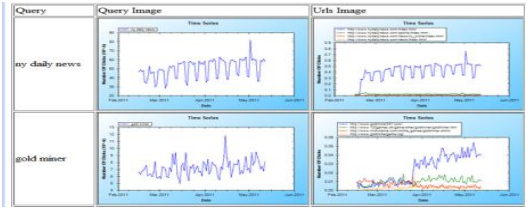
Temporal Retrieval Models

- Ranking algorithms look only at a single snapshot of a page
- Leveraging content change on a page
  - Pages have different *rates of change* (i.e., a temporal prior)
  - Terms have different *longevity* on a page
  - Results



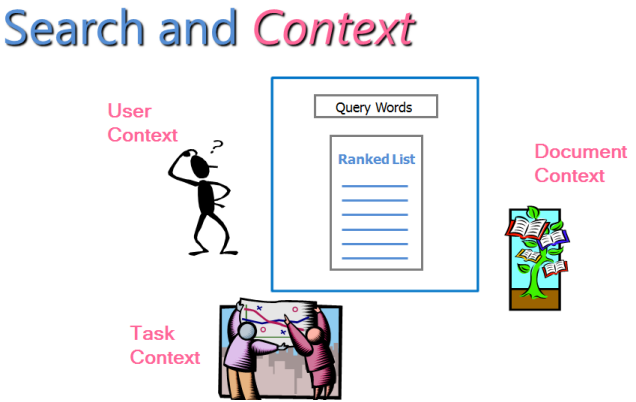
- Leveraging time-series modeling of user interactions
  - Model Query and URL clicks as time-series
  - Predict clicks at any point in time
  - Results

	Avg. Pearson
Ranker with text features	0.012
Ranker with text features and avg. click as feature	0.150
Ranker with text features and time series modeling prediction as feature	0.300



Summary

- Understanding a searcher's intent is difficult
- Augmenting text with context important
  - Who, what, where, when?
  - Why are you asking?
- Think outside the search box !!!



Thanks!

- Questions?
- More info: http://research.microsoft.com/~sdumais

- References
  - Long-term models
    - Teevan et al., SIGIR 2005. Personalizing search via automated analysis of interests and activities.
    - Teevan et al., TOCHI 2010. Potential for personalization.
  - 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.
  - Location
    - Mei and Church, WSDM 2008. Entropy of search logs: How hard is search?
    - Radlinski et al., SIGIR 2011. Inferring and using location metadata to personalize web search.
  - Time
    - Adar et al., WSDM 2009. The web changes everything: Understanding the dynamics of web content.
    - 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., in preparation. Temporally-aware ranking.
  - General
    - Dumais, UMAP 2009. Thinking outside the search box.
    - Pedersen, SIGIR 2010. Query understanding at Bing.

Microsoft Research
FacultySummit

FUTURE/WORLD 2011 2031

Microsoft Research
FacultySummit

FUTURE/WORLD

2011 — 2031