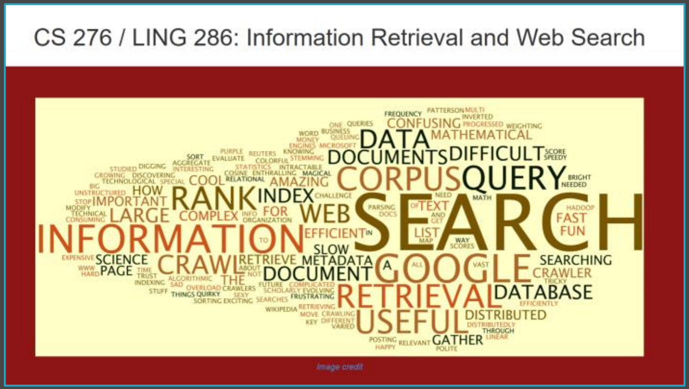


PERSONALIZED SEARCH

May 23, 2019

Susan Dumais, Microsoft Research

My Background

- Mathematics & Psychology
- Bell Labs
  - Vocabulary mismatch
    - Rich aliasing, adaptive indexing, latent semantic indexing (LSI)
    - Modeling vocabulary acquisition
- Microsoft Research
  - Text classification (e.g., spam filter)
  - Context & search (e.g., re-finding, personalization, task support)
  - Personal web of information
- Evolution of search
- Better together

My Background

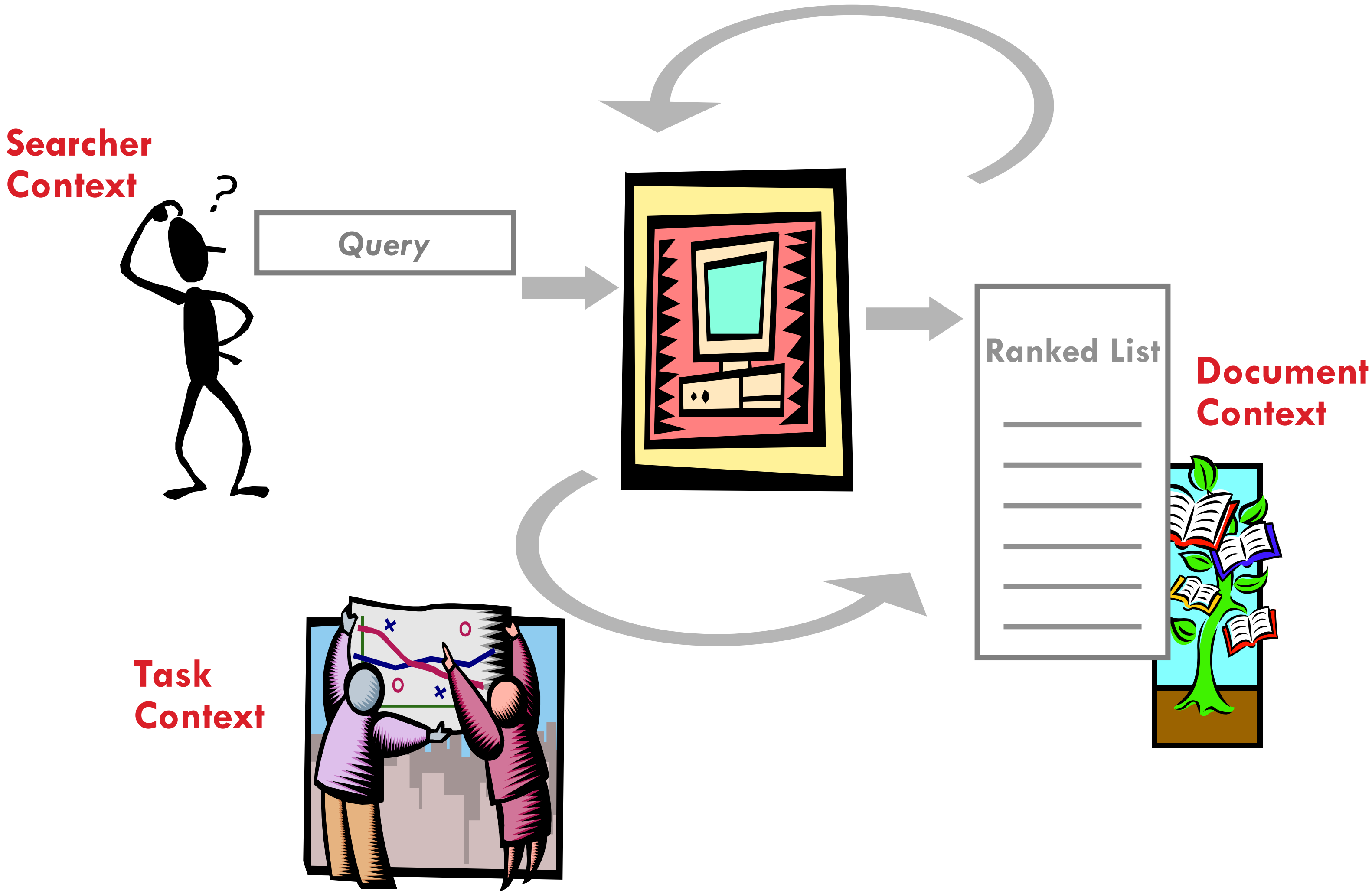
- Mathematics & Psychology
- Bell Labs
  - Vocabulary mismatch
    - Rich aliasing, adaptive indexing, latent semantic indexing (LSI)
    - Modeling vocabulary acquisition
- Microsoft Research
  - Text classification (e.g., spam filter)
  - Context & search (e.g., re-finding, personalization, task support)
  - Personal web of information
- Evolution of search
- Better together: *If search doesn't work for people, it doesn't work!*

Overview

- Personalized search perspectives
  - Context in search
  - Potential for personalization framework
- Examples
  - Personal navigation
  - Client-side personalization
  - Short- and long-term models
  - Spatio-temporal contexts
  - Personal crowds
- Challenges and new directions

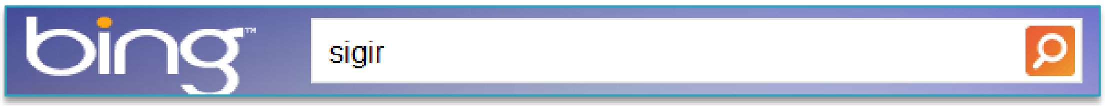
Context & Potential for Personalization

Search in Context



Context Improves Query Understanding

- Queries are difficult to interpret in isolation



- 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 … 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)

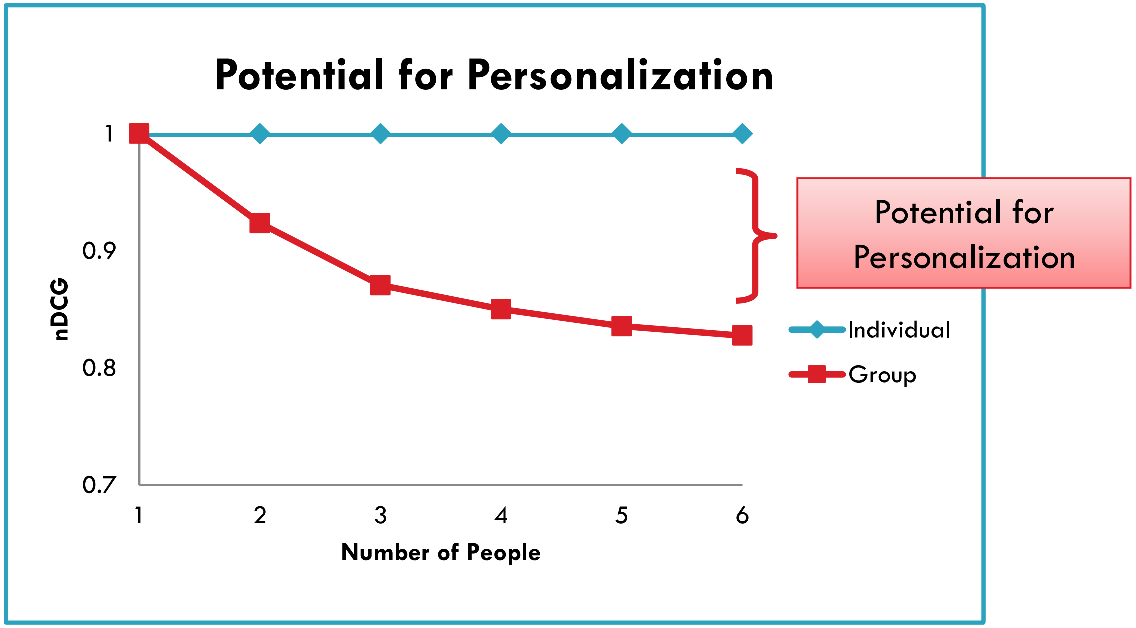
  Location: (*SIGIR* | at SIGIR conference) vs. (*SIGIR* | in Washington DC)

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

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

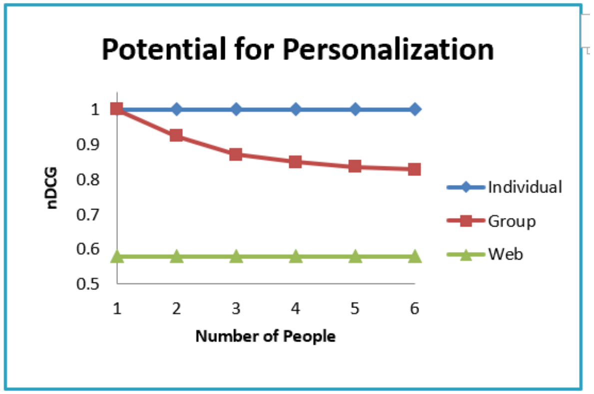
Potential For Personalization

- A single ranking for everyone limits search quality
- Quantify the variation in relevance for the same query across different individuals



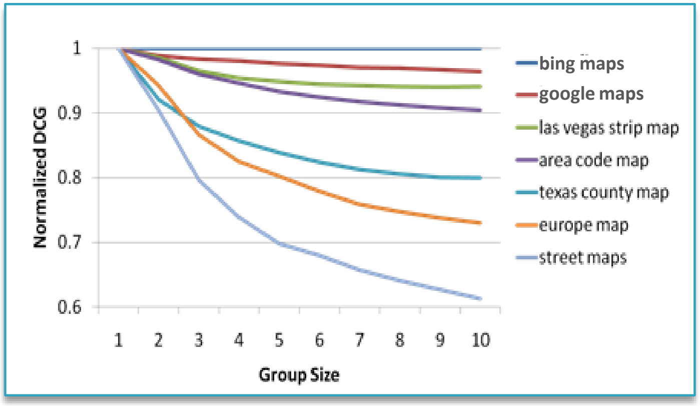
Potential For Personalization

- A single ranking for everyone limits search quality
- Quantify the variation in relevance for the same query across different individuals
- Different ways to measure individual relevance
  - Explicit judgments from different people for the same query
  - Implicit judgments from click entropy or content analysis
- Personalization can lead to large improvements
  - Study with explicit judgments
  - 46% improvements for core ranking
  - 70% improvements with personalization

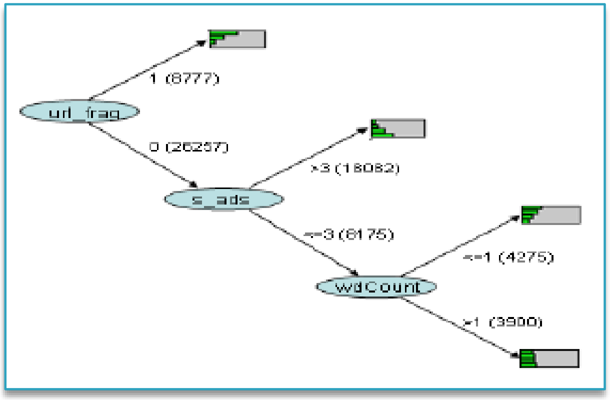


Potential For Personalization

- Not all queries have high potential for personalization
  - E.g., *new york times* vs. *sigir*
  - E.g., * *maps*

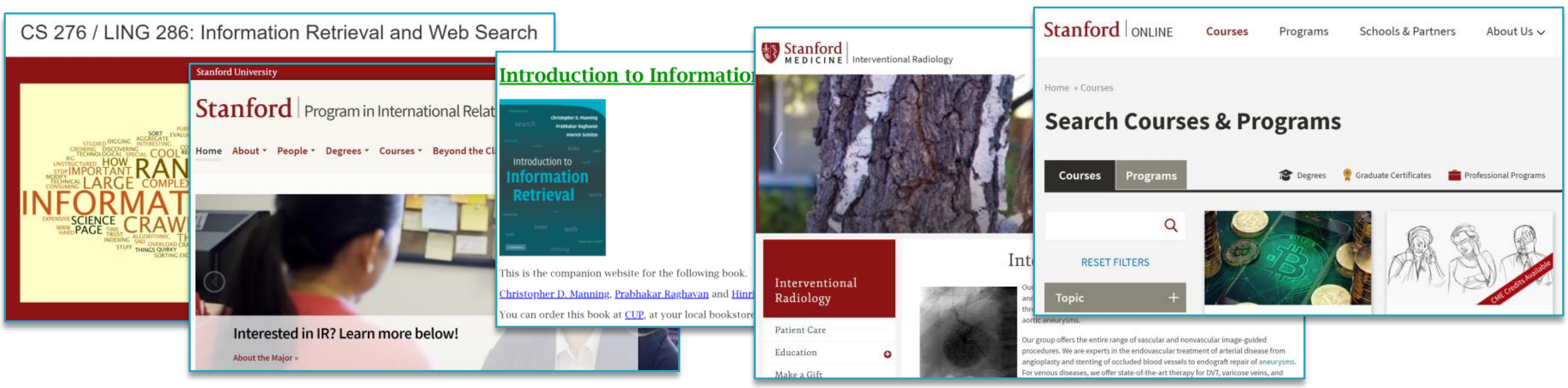


- Learn when to personalize



Potential for Personalization

- Query: *Stanford IR course*
- What is the "potential for personalization"?



- How can you identify different intents?
  - Past behavior - Current session, Longer history of actions and preferences
  - Contextual metadata – Location, Time, Device, etc.

User Models

- Constructing user models
  - Sources of evidence
    - Content: Queries, content of web pages, desktop index, etc.
    - Behavior: Explicit feedback, implicit feedback, visited web pages
    - 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

PNav

PSearch

Short/Long

Examples Methods & Applications

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., *new york times* -> www.nytimes.com
  - Consistent intent across individuals
  - Identified via low click entropy, anchor text
- "Personal navigational" queries
  - Different intents across individuals … but consistently the same intent for an individual
    - *SIGIR* (for Dumais) -> www.sigir.org
    - *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

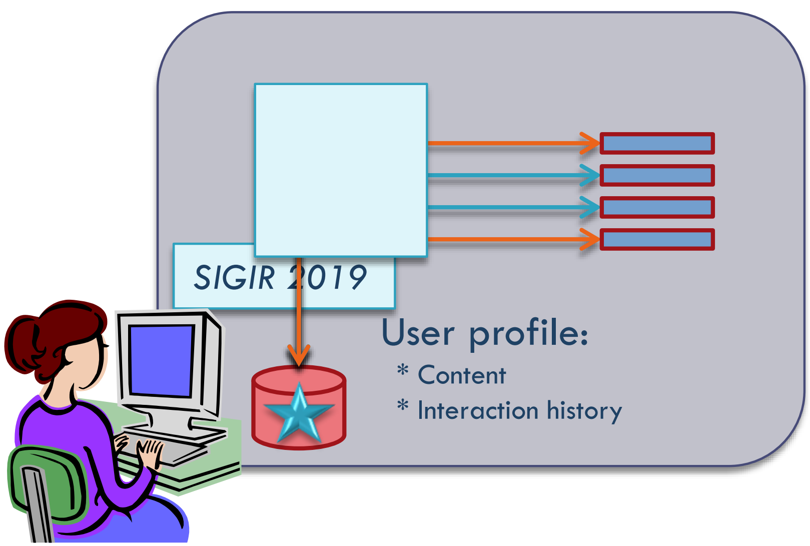
- Large-scale log analysis (offline)
  - Identifying personal navigation queries
    - Use consistency of queries & clicks <u>within an individual</u>
    - Specifically, the last two times a person issued the query, did they have a unique click on same result?
  - Coverage and prediction
    - Many such queries: ~12% of queries
    - Prediction accuracy high: ~95% accuracy
    - High coverage, low risk personalization
- A/B *in situ* evaluation (online)
  - Confirmed benefits

Teevan et al., *SIGIR 2005, ToCHI 2010*

Example 2: PSearch

- Rich client-side model of a person's interests
  - Model: Content from desktop search index & Interaction history

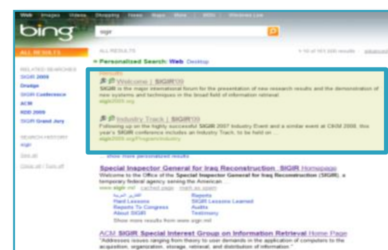
    Rich and constantly evolving user model
  - Client-side re-ranking of web search results using model
  - Good privacy (only the query is sent to server)
    - But, limited portability, and use of community

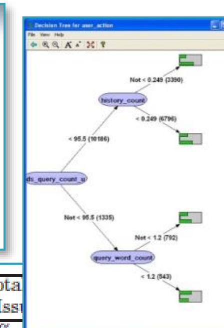


PSearch Details

- Personalized ranking model
  - Score: Weighted combination of personal and global web features
    - $Score(result_i) = \alpha\, WebScore(result_i) + (1-\alpha)\, PersonalScore(result_i)$
  - 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
    - CTR 28% higher, for personalized results

      CTR 74% higher, when personal evidence is strong
    - Learned model for when to personalize





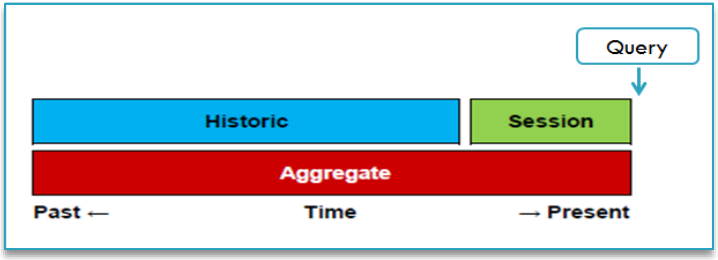
		Personalized Result Clicks	% of tota Queries Issu
Web results		4.3%	36.1%
Personalized		5.5%	63.9%
Items matched	1-5	4.2%	22.4%
	6–10	5.2%	8.5%
	11–50	6.0%	17.2%
	51–100	5.6%	5.5%
	100+	7.5%	10.3%

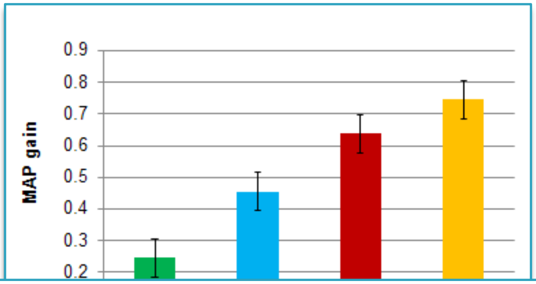
Example 3: Short + Long

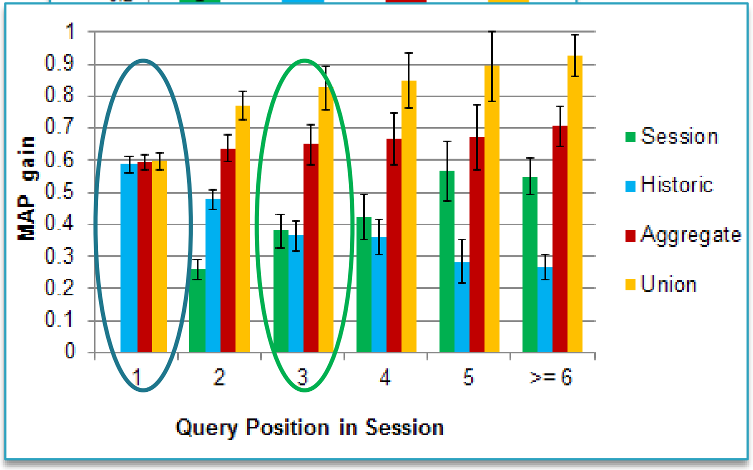
- Long-term preferences and interests
  - Content: Language models, topic models, etc.
  - Behavior: Specific queries, URLs
- Short-term context or task
  - 60% of search session have multiple queries
  - Actions within current session (Q, click, topic)
    - (Q=*sigir* | *information retrieval* vs. *iraq reconstruction*)
    - (Q=*Stanford IR course* | *CS276* vs. *intnl relations* vs. *radiology*)
    - (Q=*ego* | *id* vs. *eldorado gold corporation* vs. *dangerously in love*)
- Personalized ranking model combines both

Short + Long Details

- User model (temporal extent)
  - Session, Historical, Combinations
  - Temporal weighting
- Large-scale log analysis
- Which sources are important?
  - Session (short-term): +25%
  - Historic (long-term): +45%
  - Combinations: +65-75%
- What happens within a session?
  - 1st query, can only use historical
  - By 3rd query, short-term features more important than long-term







Atypical Sessions

- Example user model

55% Football ("nfl","philadelphia eagles","mark sanchez")
14% Boxing ("espn boxing","mickey garcia","hbo boxing")
9% Television ("modern familiy","dexter 8","tv guide")
6% Travel ("rome hotels","tripadvisor seattle","rome pasta")
5% Hockey ("elmira pioneers","umass lax","necbl")

New Session 1:
Boxing ("soto vs ortiz hbo")
Boxing ("humberto soto")

Typical

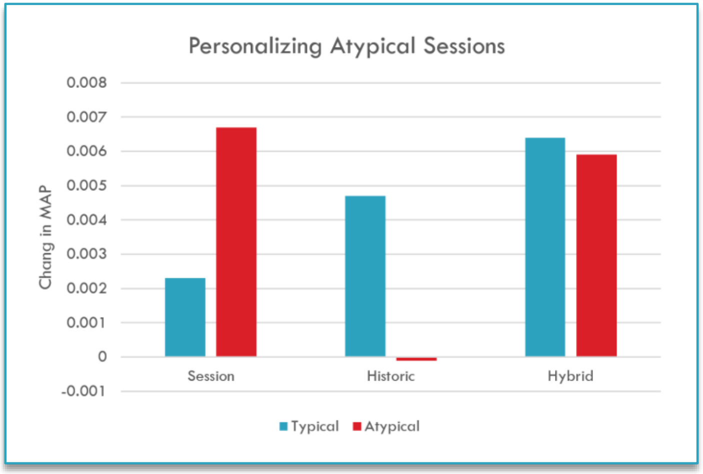
New Session 2:
Dentistry ("root canal")
Dentistry ("dental implant")
Healthcare ("dental implant recovery")

Atypical

- ~6% of sessions are atypical
  - Common topics: Medical (49%), Computers (24%)
  - Tend to be more complex, and have poorer quality results
  - What you "need" to do vs. what you "choose" to do

Atypical Sessions Details

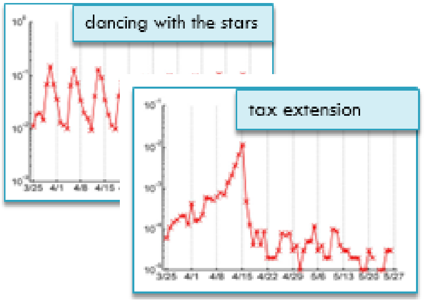
- Learn model to identify atypical sessions
  - Logistic regressions classifier
- Apply different personalization models for them
  - If typical, use long-term user model
  - If atypical, use short-term session user model
- Change in precision by typicality of session



Elsas & Dumais, *WSDM 2010*
Radinski et al., *TOIS 2013*

Example: Temporal Dynamics

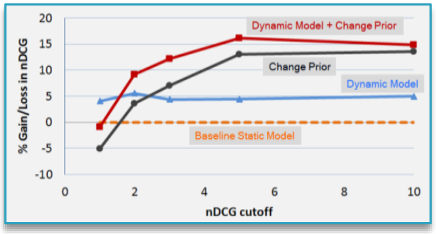
- Queries are not uniformly distributed over time
  - Often triggered by events in the world



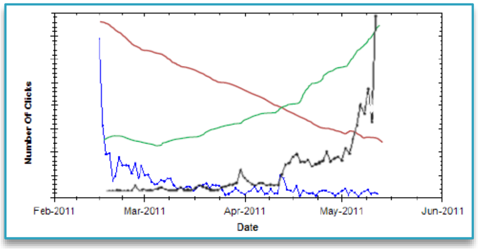
- What's relevant changes over time
  - E.g., *US Open* … in 2019 vs. in 2018
  - E.g., *US Open 2019* … in May (golf) vs. in Sept (tennis)
  - E.g., *US Tennis Open 2019* …
    - Before event: Schedules and tickets, e.g., stubhub
    - During event: Real-time scores or broadcast, e.g., espn
    - After event: General sites, e.g., wikipedia, usta

Temporal Dynamics Details

- Develop time-aware retrieval models
- Model <u>content</u> 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 <u>user interactions</u> 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

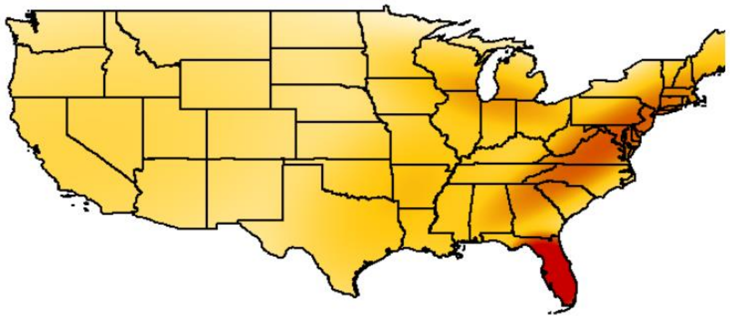


Example: Location Context

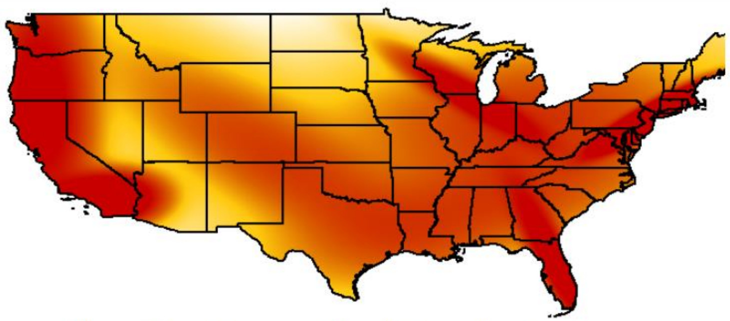
- What's relevant to a query varies by location
  - E.g., *football* or *jumper or chips* [in US vs. UK]
  - E.g., *library or zoo or current time* [at state- or city-level]
  - E.g., *starbucks or pizza* [at finer granularity]
- Data: query, URL, location
- Geographic distribution of each URL, query
  - $P(location = X \mid URL)$, estimate this using a mixture of Gaussians
  - $P(location \mid query)$, estimate this using a mixture of Gaussians
  - Background model

Location Context Details

- Location interest model for Q: *smh*

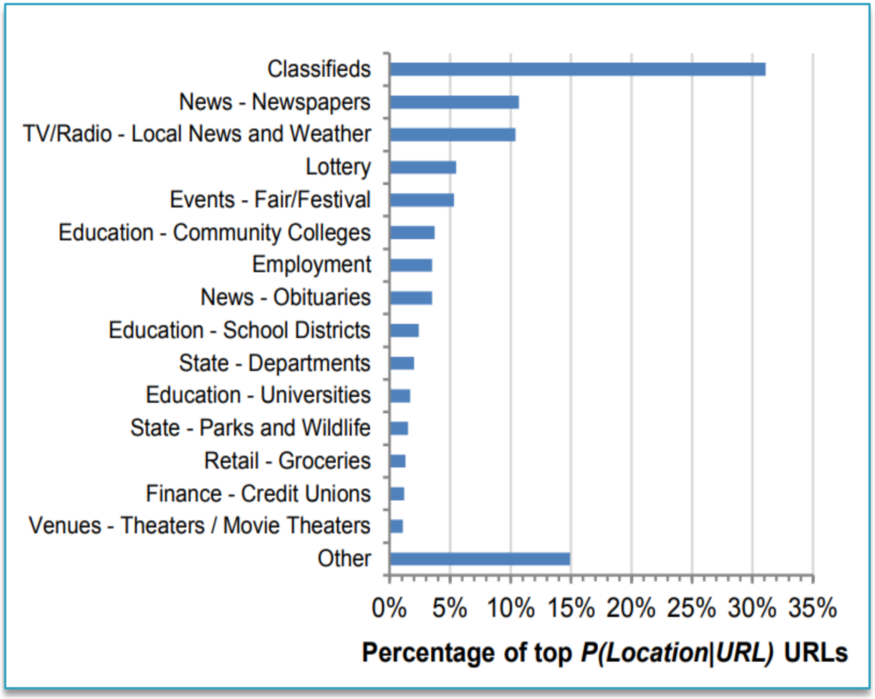


(a) Sarasota Memorial Health, http://smh.com/



(b) Sydney Morning Herald, http://smh.com.au/

- Topics w/ the most location-centric URLs



Location Context Details

- Learn to re-rank using location features
- Important features
  - Original ranking
  - $P(URL \mid searcher\ location)$
  - $KL\ Div(URL\ model, background\ model)$
- Query: *rta bus schedule*

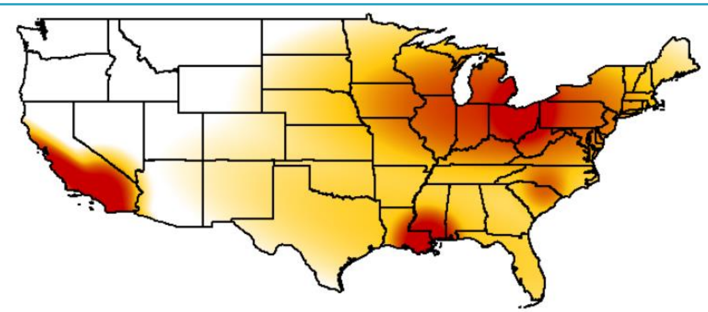


Figure 4. Query location model for the query [*rta bus schedule*]. We see three peaks: in Ohio, Louisiana, and California.

Challenges & Opportunities

Challenges in Personalization

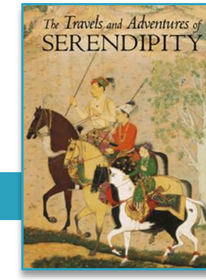
- User-centered
  - Privacy
  - Serendipity and novelty
  - Transparency and control
- Systems-centered
  - Optimization
    - Storage, run-time, caching, etc.
  - Evaluation
    - Measurement, experimentation

Privacy

- Profile and content need to be in the same place
- Local profile (e.g., PSearch)
  - Private, only query sent to server
  - Device specific, inefficient, no community learning
- Cloud profile (e.g., Web search)
  - Need transparency and control over what's stored
- Other approaches
  - Public/semi-public profiles (e.g., tweets, FB status, blogs, papers)
  - Light weight profiles (e.g., queries in a session)
  - Matching to a group cohort vs. an individual

André et al., *CHI 2009*, *C&C 2009*

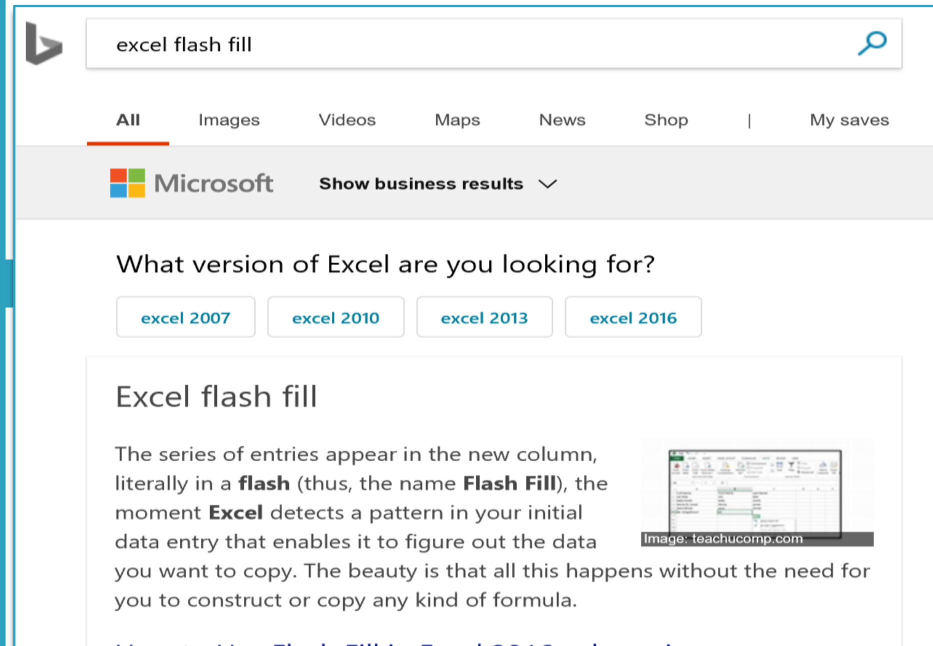
Serendipity and Novelty

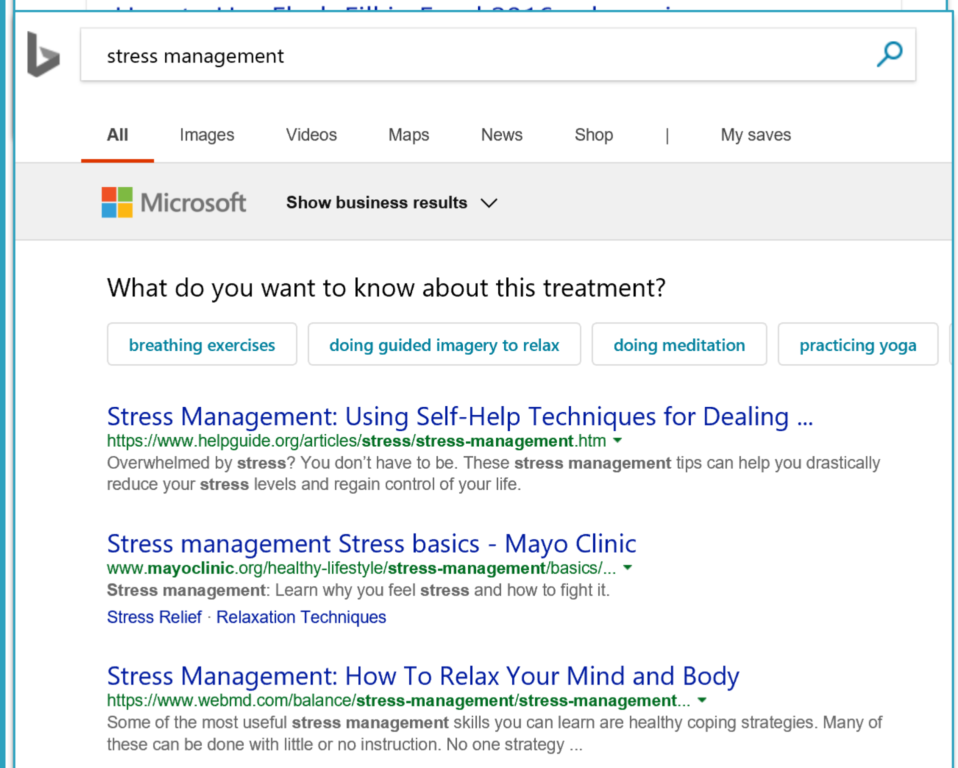


- Does personalization mean the end of serendipity?
  - … Actually, it can improve it!
- Experiment on *Relevance vs. Interestingness*
  - Personalization finds more <u>relevant</u> results
  - Personalization also finds more <u>interesting</u> results
    - Even when interesting results were not relevant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

Bing

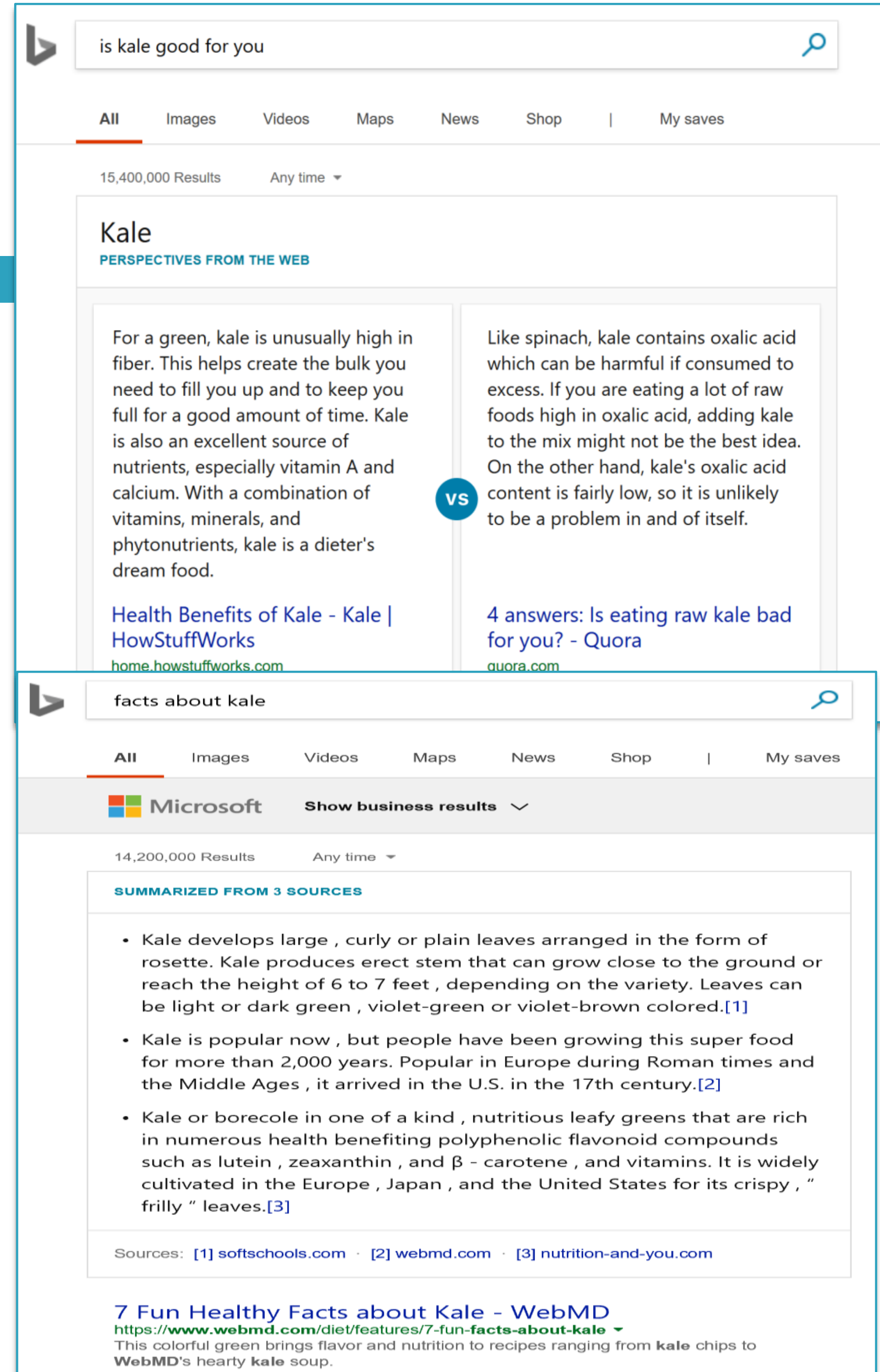
Perspectives



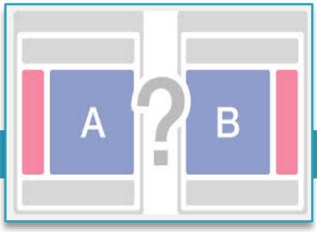


Bing

Perspectives



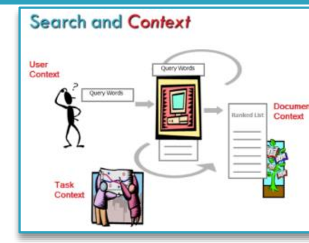
Evaluation



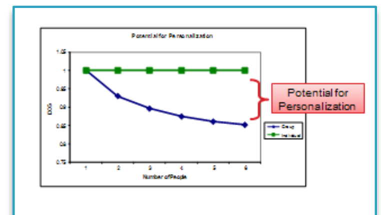
- External judges, e.g., assessors
  - Lack diversity of intents and realistic context
  - Crowdsourcing can help some
- Actual searchers are the "judges"
  - Offline
    - Labels from explicit judgments or implicit behavior (log analysis)
    - Allows safe exploration of many different alternatives
  - Online (A/B experiments)
    - Explicit judgments: Nice, but annoying and may change behavior
    - Implicit judgments: Scalable and natural, but can be very noisy
- Linking implicit actions and explicit judgments

Personalized Search Recap

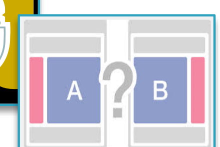
- Queries difficult to interpret in isolation
  - Augmenting query with context critical



- Large potential for improving search via personalization



- Examples
  - PNav, PSearch, Short/Long, Crowd
- Challenges
  - Privacy, transparency, serendipity
  - Evaluation, system optimization



- Personalization/contextualization prevalent today, and increasingly so in mobile and proactive scenarios

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, Sarah Tyler, Alex Kotov, Paul André, Carsten Eickhoff, Peter Organisciak

References

- André et al., CHI 2009. *From x-rays to silly putty via Uranus: Serendipity and its role in Web search.*
- Bennett et al., SIGIR 2011. *Inferring and Using Location Metadata to Personalize Web Search.*
- Bennett et al., SIGIR 2012. *Modeling the impact of short- & long-term behavior on search personalization.*
- Dumais et al., Ways of Knowing 2014. *Understanding behavior through log data and analysis.*
- Eickhoff et al., WSDM 2013. *Personalizing atypical search sessions.*
- Elsas et al., WSDM 2010. *Leveraging temporal dynamics of document content in relevance ranking.*
- Organisciak et al., HCOMP 2015. *A crowd of your own: Crowdsourcing for on-demand personalization.*
- Radiniski et al., TOIS 2013. Behavioral dynamics on the web: Learning, modeling and predicting
- 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.*