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

Susan Dumais
Principal Researcher
Microsoft Research
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

User Context

Task Context

Query Words

Ranked List

Document Context
Using Context to Improve Query Understanding

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

- 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) vs. (SIGIR | Stuart Bowen Jr.)

**Previous Actions:** (SIGIR | Information Retrieval vs. (SIGIR | Information Retrieval)

**Location:** (SIGIR | Beijing | Spring) vs. (SIGIR | New York)

**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

User profile:
- * Content
- * Interaction history
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
  \[
  \text{Score} (\text{result}_i) = \alpha \text{PersonalScore} (\text{result}_i) + (1 - \alpha) \text{WebScore} (\text{result}_i)
  \]

**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

<table>
<thead>
<tr>
<th>Personalized Result Clicks</th>
<th>% of total Queries Issued</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web results</td>
<td>4.3%</td>
</tr>
<tr>
<td>1-5</td>
<td>9.0%</td>
</tr>
<tr>
<td>6-10</td>
<td>4.2%</td>
</tr>
<tr>
<td>11-50</td>
<td>5.2%</td>
</tr>
<tr>
<td>51-100</td>
<td>6.0%</td>
</tr>
<tr>
<td>100+</td>
<td>5.5%</td>
</tr>
</tbody>
</table>
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

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

<table>
<thead>
<tr>
<th>Context source</th>
<th>Accuracy ($F_1$)</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Query</td>
<td>Context</td>
</tr>
<tr>
<td>None (i.e., current query only)</td>
<td>0.39</td>
<td>–</td>
</tr>
<tr>
<td>Queries (i.e., all previous queries)</td>
<td>0.39</td>
<td>0.42</td>
</tr>
<tr>
<td>Queries + SERPClicks (i.e., all previous queries / result clicks)</td>
<td>0.39</td>
<td>0.46</td>
</tr>
<tr>
<td>Queries + SERPClicks + NavTrails (i.e., all previous actions)</td>
<td>0.39</td>
<td>0.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context source</th>
<th>Percentage of queries best between models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Query</td>
</tr>
<tr>
<td>Queries (i.e., all previous queries)</td>
<td>25%</td>
</tr>
<tr>
<td>Queries + SERPClicks (i.e., all previous queries / result clicks)</td>
<td>30%</td>
</tr>
<tr>
<td>Queries + SERPClicks + NavTrails (i.e., all previous actions)</td>
<td>34%</td>
</tr>
</tbody>
</table>
Location

- How much does knowing location help search?
  - Search: $H(\text{URL} \mid \text{Query}) = 2.8$
  - Search & Location: $H(\text{URL} \mid \text{Query, IP}) = 1.2$
- 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 changes
  - 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
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](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**

- **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**
  - Elsas and Dumais, WSDM 2010. Leveraging temporal dynamics of document content in relevance ranking.
  - Radinsky et al., in preparation. Temporally-aware ranking.

- **General**
  - Dumais, UMAP 2009. Thinking outside the search box.
  - Pedersen, SIGIR 2010. Query understanding at Bing.
© 2011 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.

MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.