Using Context to Support Searchers in Searching

Susan Dumais
Microsoft Research

http://research.microsoft.com/~sdumais
Using **Context** to Support Searchers

**User Context**

**Query Words**

**Ranked List**

**Document Context**

**Task/Use Context**
Web Info through the Years

What’s available

- Number of pages indexed
  - 7/94 Lycos
  - 95 - $10^6$ millions
  - 97 - $10^7$
  - 98 - $10^8$
  - 01 - $10^9$ billions
  - 05 - $10^{10}$ ...

- Types of content
  - Web pages, newsgroups
  - Images, videos, maps
  - News, blogs, spaces
  - Shopping, local, desktop
  - Books, papers
  - Health, finance, travel ...

How it’s accessed

Search the Web:
Some Support for Searchers

- The search box
- Spelling suggestions
- Query suggestions
- Advanced search operators and options (e.g., "", +/-, site:, language:, filetype:, intitle:)
- Richer snippets
- But, we can do better … using context
Key Contexts

- Users:
  - Individual, group (topic, time, location, etc.)
  - Short-term or long-term models
  - Explicit or implicit capture

- Documents/Domains:
  - Document-level metadata, usage/change patterns
  - Relations among documents

- Tasks/Uses:
  - Information goal - Navigational, fact-finding, informational, monitoring, research, learning, social, etc.
  - Physical setting - Device, location, time, etc.
Using Contexts

- **Identify:**
  - What context(s) are of interest?

- **Accommodate:**
  - What do we do differently for different contexts?
  - Outcome (Q|context) >> Outcome (Q)

- **Influence points within the search process**
  - Articulating the information need
    - Initial query, subsequent interaction/dialog
  - Selecting and/or ranking content
  - Presenting results
  - Using and sharing results
Context in Action

Research prototypes: provide insights about algorithmic, user experience, and policy challenges

- User Contexts:
  - Finding and Re-Finding (Stuff I’ve Seen)
  - Personalized Search (PSearch)
  - Novelty in News (NewsJunkie)

- Document/Domain Contexts:
  - Metadata and search (Phlat)
  - Visualizing patterns in results (GridViz)

- Task/Use Contexts:
  - Pages as context (Community Bar, IQ)
  - Richer collections as context (NewsJunkie, PSearch)
  - Working, understanding, sharing (SearchTogether, InkSeine)
**SIS: Stuff I’ve Seen**

- Unified index of *stuff you’ve seen*
  - Many info silos (e.g., files, email, calendar, contacts, web pages, rss, im)
  - Unified index, not storage
  - Index of content and metadata (e.g., time, author, title, size, access)
  - Re-finding vs. finding

**Vista Desktop Search** (and Live Toolbar)

Also, Spotlight, GDS, X1, ...
## SIS Demo

### Stuff I’ve Seen

<table>
<thead>
<tr>
<th>Document</th>
<th>Date</th>
<th>Rank</th>
<th>Path</th>
<th>Author</th>
<th>Mail To</th>
</tr>
</thead>
<tbody>
<tr>
<td>(All) (200)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web Pages (77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outlook (72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Files (51)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**200 rows returned**

- **Briefing:** AP/Sue Dumais 5/6/2004 at 4:30 p.m.
  - Date: 5/6/2004 11:27...
  - Path: personal folders/media
  - Author: Kim Davis
  - Mail To: Susan Dumais

- **Final Recap:** The Economist Campus Visit April 15...
  - Date: 4/20/2004 6:21...
  - Path: personal folders/media
  - Author: Kristen Birkeland
  - Mail To: Karen Redetzki; Suzan DelBene; John

- **Last 7 days**
  - Microsoft: Longhorn to arrive in 2005 - News - ZD...
    - Date: 3/26/2004 10:22...
    - Path: temporary internet files\content.ie5\vty7ouub
  - News: All about Longhorn
    - Date: 3/26/2004 10:2...
    - Path: temporary internet files\content.ie5\0xubghir
  - X1 instantly searches files & email. For Outlook, Ou...
    - Date: 3/24/2004 10:11...
    - Path: temporary internet files\content.ie5\v73fh0w
  - Enhanced Microsoft: Exhibit Offers Sneak Peek at ...
    - Date: 3/23/2004 12:22...
    - Path: temporary internet files\content.ie5\v5nshv7a
  - Prefaceprogram
    - Date: 3/22/2004 5:01...
    - Path: personal folders\hit-naacl04
  - prefecparent
    - Date: 3/22/2004 4:19...
    - Path: d:\personal\papers\review\hit-naacl04
  - HLT/NAACL Preface
    - Date: 3/22/2004 8:18...
    - Path: personal folders\hit-naacl04
  - The Future of Information Filtering
    - Date: 3/20/2004 12:04...
    - Path: temporary internet files\content.ie5\szfn6cx
  - Slashdot | Google buys Pyra Labs
    - Date: 3/17/2004 3:22...
    - Path: temporary internet files\content.ie5\vmko3hds5
  - Recap: Steven Levy/Newsweek on IQ
    - Date: 3/15/2004 6:27...
    - Path: temporary internet files\content.ie5\vsvv4gsk
  - Homepage for HLT-NAACL 2003
    - Date: 3/15/2004 5:34...
    - Path: temporary internet files\content.ie5\ep381kvuy
  - Pedro Domingos
    - Date: 3/10/2004 4:00...
    - Path: temporary internet files\content.ie5\zvbfu5t
  - JEP: Multimedia Features and Information Retrieval
    - Date: 3/8/2004 3:35...
    - Path: temporary internet files\content.ie5\vkkkr15h
  - Bates’ Bibliography: Information Seeking, Indexing...
    - Date: 3/8/2004 3:35...
    - Path: temporary internet files\content.ie5\ep381kvuy

- **Last 30 days**

- **Older than 30 days**

ACL/HLT – June 18, 2008
SIS Usage Experiences

Internal deployment
- ~3000 internal Microsoft users
- Analyzed: Free-form feedback, Questionnaires, Structured interviews, Log analysis (characteristics of interaction), UI expts, Lab expts

Personal store characteristics
- 5k - 500k items

Query characteristics
- Short queries (1.6 words)
- Few advanced operators or fielded search in query box (~7%)
- Many advanced operators and query iteration in UI (48%)
  - Filters (type, date); modify query; re-sort results

| Susan's (Laptop) World |   |   |
|------------------------|---|
| **Type** | **N** | **Size** |
| Web | 3k | 0.2 Gb |
| Files | 28k | 23.0 GB |
| Mail | 60k | 2.2 Gb |
| **Total** | **91k items** | **25.4 Gb** |
| **Index** |   | **190 Mb** |
|           |   | +1.5 Mb/week |
Importance of people, time, and memory

- **People**
  - 25% of queries contained names
  - People in roles (to:, from:) vs. people as entities in text

- **Time**
  - Age of items opened
    - 5% today; 21% last week
    - 50% of the cases in 36 days
    - Web (11); Mail (36); Files (55)
  - **Date** most common sort field, even when Rank was the default
    - Support for episodic memory

- Few searches for “best” topical match ... many other criteria

Log(Freq) = -0.68 * log(DaysSinceSeen) + 2.0
Observations about unified access

- Metadata quality is variable
  - Email: rich, pretty clean
  - Web: little, available to application
  - Files: some, but often wrong

- Memory depends on abstractions
  - “Useful date” is dependent on the object!
    - Appointment, when it happens
    - File, when it is changed
    - Email and Web, when it is seen
  - “People” attribute vs. contains
    - To, From, Cc, Attendee, Author, Artist
Ranked list vs. Metadata (for personal content)

Why Rich Metadata?

• People remember many attributes in re-finding
  - Often: time, people, file type, etc.
  - Seldom: *only* general overall topic
• Rich client-side interface
  - Support fast iteration/refinement
  - Fast filter-sort-scroll vs. next-next-next-next
Re-finding on the Web

- 50-80% URL visits are revisits
- 30-40% of queries are re-finding queries

Table 1. A classification of different query types.

<table>
<thead>
<tr>
<th>All queries: 13,060 queries (100%)</th>
<th>Overlapping Click Queries – 5072 queries (39%)</th>
<th>Equal Click Queries – 3777 (29%)</th>
<th>Some Common Clicks 1295 (10%)</th>
<th>No Common Clicks 7988 (61%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal Query Queries 4256 (33%)</td>
<td>Navigational Queries 3100 (24%)</td>
<td>Single Identical Click 3737 (29%)</td>
<td>Multiple Identical Clicks 40 (&lt; 1%)</td>
<td></td>
</tr>
<tr>
<td>Different Query 8804 (67%)</td>
<td></td>
<td>637 (5%)</td>
<td>4 (&lt; 1%)</td>
<td>7503 (57%)</td>
</tr>
</tbody>
</table>

Teevan et al., SIGIR 2007
Phlat: Search and Metadata

- Shell for WDS; publically available
- Features:
  - Search / Browse (faceted metadata)
  - Unified Tagging
  - In-Context Search
Phlat: Faceted metadata

- Tight coupling of search and browse
- Q → Results &
  - Associated metadata w/ query previews
  - 5 default properties to filter on (extensible)
  - Includes tags
- Property filters integrated with query
  - Query = words and/or properties
  - No stuck filters
- Search == Browse
Phlat: Tagging

- Apply a **single set** of user-generated tags to **all content** (e.g., files, email, web, rss, etc.)

- Tagging interaction
  - Tag widget or drag-to-tag

- Tag structure
  - *Allow* but do not *require* hierarchy

- Tag implementation
  - Tags directly associated with files as NTFS or MAPI properties
Phat: In-Context Search

- Selecting a result ...
- Linked view to show associated tags
- Rich actions
  - Open, drag-drop, etc.
- Pivot on metadata
  - “Sideways search”
- Refine or replace query
Phlat shell for Windows Desktop Search

- Tight coupling of searching/browsing
- Rich faceted metadata support
  - Including unified tagging across data types
- In-context search and actions

Download: http://research.microsoft.com/adapt/phlat
Web Search using Metadata

Many queries include implicit metadata:
- portrait of barak obama
- recent news about midwest floods
- good painters near redmond
- starbucks near me
- overview of high blood pressure

Limited support for users to articulate this
Search in Context

- Search is not the end goal ...
- Support information access in the context of ongoing activities (e.g., writing talk, finding out about, planning trip, buying, monitoring, etc.)

- Search always available
- Search from within apps (keywords, regions, full doc)
- Show results within app
- Maintains “flow” (Csikszentmihalyi)
- Can improve relevance
Documents as (a simple) Context

Proactive “query” specification depending on current document content and activities

- **Recommendations**
  - People who bought this also bought …

- **Contextual Ads**
  - Ads relevant to page

- **Community Bar**
  - Notes, Chat, Tags, Inlinks, Queries

- **Implicit Queries (IQ)**
  - Also Y!Q, Watson, Rememberance Agent
Document Contexts (Implicit Query, IQ)

- Proactively find info related to item being read/created
  - Quick links
  - Related content

- Challenges
  - Relevance, fine
  - When to show? (useful)
  - How to show? (peripheral awareness)

Quick links for People and Subject.

Background search on top k terms, based on user’s index — Score = \( \frac{tf_{doc}}{\log(tf_{corpus}+1)} \)

Top matches for this Implicit Query (IQ).
**PSearch: Personalized Search**  
(Even Richer Context)

- Today: People get the same results, independent of current session, previous search history, etc.
- PSearch: Uses rich client-side info to personalize results

- Building a user profile
- Personalized ranking
- When to personalize?
- How to personalize display?

---

**ACM SIGIR** Special Interest Group on Information Retrieval Home Page
Welcome to the ACM SIGIR Web site … SIGIR thanks Doug Oard, Bill Hersh, David Carmel, Noriko Kando, Diane Kelly… Get ready for SIGIR 2008!
[sigir.org](http://sigir.org)
Building a User Profile

- Type of information:
  - Explicit: Judgments, categories
  - Content: Past queries, web pages, desktop
  - Behavior: Visited pages, dwell time

- Time frame: Short term, long term

- Who: Individual, group

- Where the profile resides:
  - Local: Richer profile, improved privacy
  - Server: Richer communities, portability
Personalized Ranking

- Personal Rank = \( f(\text{Cont}, \text{Beh}, \text{Web}) \)
  - Pers_Content Match: \( \text{sim}(\text{result}, \text{user_content_profile}) \)
  - Pers_Behavior Match: \( \text{visited URLs} \)
  - Web Match: \( \text{web rank} \)
When to Personalize?

- Personalization works well for some queries, ... but not for others
- Framework for understanding when to personalize
  - Personal ranking
    - Personal relevance (explicit or implicit)
  - Group ranking
    - Decreases as you add more people
  - Gap is “potential for personalization (p4p)”
More Personalized Search

- PSearch - rich long-term context; single individual
- Short-term session/task context
  - Session analysis
  - Query: ACL, ambiguous in isolation
    - Natural language ... summarization ... ACL
    - Knee surgery ... orthopedic surgeon ... ACL
- Groups of similar people
  - Groups: Location, demographics, interests, behavior, etc.
  - Mei & Church (2008)
    - H(URL) = 22.4
    - Search: H(URL|Q) = 2.8
    - Personalization: H(URL|Q, IP) = 1.2
- Many models ... smooth individual, group, global models
Beyond Search - Gathering Info

- Support for more than retrieving documents
  - Retrieve -> Analyze -> Use
- Lightweight scratchpad or workspace support
  - Iterative and evolving nature of search
  - Resuming at a later time or on other device
  - Sharing with others
Beyond Search - Sharing & Collaborating

- **SearchTogether**
  - Collaborative web search prototype
  - Sync. or async. sharing w/ others or self

- Collaborative search tasks
  - E.g., Planning travel, purchases, events; understanding medical info; researching joint project or report

- Today little support
  - Email links, instant messaging, phone

- SearchTogether adds support for
  - Awareness (history, metadata)
  - Coordination (IM, recommend, split)
  - Persistence (history, summaries)

Figure 1. The SearchTogether client: (a) integrating messaging, (b) query awareness, (c) current results, (d) recommendation queue, (e) (f)g search buttons, (h) page-specific metadata, (i) toolbar, (j) browser
Looking Ahead …

- Continued advances in scale of systems, diversity of resources, ranking, etc.
- Tremendous new opportunities to support searchers by
  - Understanding user intent
    - Modeling user interests and activities over time
    - Representing non-content attributes and relations
  - Supporting the search process
    - Developing interaction and presentation techniques that allow people to better express their information needs
    - Supporting understanding, using, sharing results
  - Considering search as part of richer landscape
Using Context to Support Searchers
Think Outside the IR Box(es)
Thank You!

- Questions/Comments ...


- Phlat, http://research.microsoft.com/adapt/phlat

References

- **Stuff I've Seen**

- **Phlat**

- **Memory Landmarks**

- **Personalized Search**

- **Implicit Queries**

- **Revisitation on Web**

- **InkSeine**

- **Search Together**