An Interdisciplinary Perspective on IR

Susan Dumais
Microsoft Research
Thanks!

- Salton Award Committee
- Many great colleagues
  - 1979-1997, Bell Labs/Bellcore
  - 1997-present, Microsoft Research
  - Many other collaborators ...

- Tremendous honor
- Salton number ∴
  - Michael Lesk: SMART @ Harvard (early 1960's)
  - CHI 1995 Panel: "Searching & browsing: Can we find a synergy?"
Overview

Personal reflections
- My research is interdisciplinary, at the intersection of IR and HCI
- User-centric vs. system-centric
- Empirical vs. theoretical
- Evaluation via many methods
  - Test collections, field work, prototypes, deployment experiences, lab studies, etc.

My background

Common themes
- Understanding user, domain, and task contexts

Future challenges
- Dynamics, data and more
Background

- Mathematics and Psychology
- HCI group at Bell Labs, 1979
- Introduction to IR, 1980

The problem(s) ...
- Human factors in database access
- Describing categories of objects for menu retrieval
- Verbal disagreement/Statistical semantics/Vocabulary problem

Some solutions & applications ...
- Rich aliasing / Adaptive indexing / Latent semantic indexing

Closing the loop back to psychology ...
- A solution to Plato’s problem [Psychological Review, 1997]
From Verbal Disagreement to LSI

- Observed: Mismatch between the way that people want to retrieve information from a computer and the way that systems designers describe that information
  - The trouble with UNIX
  - Command names, menu and category descriptors, keywords

- Studied: How people describe objects and operations
  - Text editing operations, systems functionality, common objects, recipes, classified ads, etc.

- Demo:

- Data:
Findings:

- Tremendous diversity in the name that people use to describe the same objects or actions (aka, “the long tail”)
  - Single keyword: 0.07 – 0.18 “repeat rate”
  - Single normative keyword: 0.16 - 0.36
  - Three aliases: 0.38 – 0.67
  - Infinite aliasing:

Interestingly, we have referred to this problem as verbal disagreement, vocabulary mismatch, statistical semantics
From Verbal Disagreement to LSI

---

**CHI 1982 Paper ... 0th CHI Conference**

**STATISTICAL SEMANTICS:** HOW CAN A COMPUTER USE WHAT PEOPLE NAME THINGS TO GUESS WHAT THINGS PEOPLE MEAN WHEN THEY NAME THINGS?

In describing items in a database, however, system designers are at a disadvantage in that they do not usually get explicit, immediate, and continuous feedback from users. Knowing how people describe common objects and shift their descriptions for audiences of different levels of sophistication may help designers build systems whose information is accessible to the widest possible audience.

---

**Photos from B. Shneiderman**
Some solutions: ... with a lot of help from our friends

Rich aliasing [Gomez et al. 1990]
- Allow alternative words for the same item
- “Natural” in the world of full-text indexing, but less so for keyword indexing or command naming

Adaptive indexing [Furnas 1985]
- Associate (failed) user queries to destination objects
- Add these queries as new entries in term-document matrix
- Quickly reduces failure rate for common requests/tasks

Latent Semantic Indexing [Dumais et al. 1988; Deerwester et al. 1990]
- Model relationships among words, using dimension reduction
- Especially useful when query and documents are short
- Baker, Borko/Bernick, Ossario (1962-1966); Kohl (SIGIR 1978, p.1)
From Verbal Disagreement to LSI

- Many applications and algorithms of LSI
  - Bell Labs directory of services, expert finding, reviewer assignment, handwritten notes, data evidence analysis, measurement of knowledge, literature-based discovery, IR & IF test collections

- Rich aliasing and Adaptive indexing in Web era
  - Full text indexing (rich aliases from authors)
  - Anchor text or Tags (rich aliases from other users)
  - Historical query-click data (adaptive indexing, with implicit measures)
Common Themes

- The last 10-20 years ... amazing time to be involved in IR
- TREC and related evaluations
  - TREC-1 in 1992
- Search is everywhere – desktop, enterprise, Web
- Web search
  - Big advances in scale, diversity of content and users, quality of results (for some tasks), etc.
- SIGIR community has a lot to be proud of
- But ... many search tasks are still quite hard
  - Need to represent and leverage richer contextual information about users, domains, and task environments in which search occurs
Web Search at 15

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, many formats
  - Health, finance, travel ...

How it’s accessed

SIGIR 2009
Support for Searchers

- The search box
- Spelling suggestions
- Query suggestions
- Auto complete
- Inline answers
- Richer snippets
- But, we can do better.

Search in the future will look nothing like today's simple search engine interfaces, [Susan Dumais] said, adding, "If in 10 years we are still using a rectangular box and a list of results, I should be fired." [Mar 7, 2007, NYTimes, John Markoff]
Search and Context

User Context

Query Words

Task/Use Context

Ranked List

Document Context
**Systems/Prototypes**
- New capabilities and experiences
- Algorithms and prototypes
- Deploy, evaluate and iterate

**Modeling Users**
- Short vs. long term
- Individual vs. group
- Implicit vs. explicit

**Using User Models**
- Stuff I’ve Seen (re-finding)
- Personalized Search
- News Junkie (novelty)
- User Behavior in Ranking
- Domain Expertise at Web-scale

**Search and Context**
- User Context
- Task/Use Context
- Document Context

**Inter-Relationships among Documents**
- Categorization and Metadata
  - Reuters, spam, landmarks, web categories...
  - Domain-specific features, time

**Interfaces and Interaction**
- Stuff I’ve Seen, Phlat, Timelines, SWISH
- Tight coupling of browsing and search

**Evaluation**
- Many methods, scales
- Individual components and their combinations

**Temporal Dynamics**
- Redundancy
User Modeling

- Modeling searcher’s interests and activities over time
  - Iterative and interactive nature of search
  - Within and across sessions

- Example applications
  - Re-finding (e.g., Stuff I’ve Seen, Web) [Dumais et al. 2003]
  - Personalization (e.g., PSearch) [Teevan et al. 2005]
  - Novelty (e.g., News Junkie) [Gabrilovich et al. 2004]
  - Domain expertise at Web-scale [White & Dumais 2009]
  - User behavior for Web ranking [Agichtein et al. 2006]

- Evaluation via explicit judgments, questionnaires, client-side instrumentation, and large-scale search logs, lab and field studies, etc.
Re-Finding on the Desktop

- Stuff I’ve Seen (SIS) [Dumais et al. 2003]:
  - Unified access to many types of info (e.g., files, email, calendar, contacts, web pages, rss, im)
  - Index of content and metadata (e.g., time, author, title, size, usage)
  - Rich UI possibilities, because it’s your stuff and client application
  - Demo:

- Analysis:
  - Deployed different versions
    - Query syntax
    - Result previews
    - Ranking defaults (time, best-match)
  - Questionnaires, Free-form feedback, Log data, Lab experiments

SIGIR 2009
Re-Finding on the Desktop

- **Research Results:**
  - Short queries
    - Few advanced operators in initial query (<10%)
    - Many advanced operators via specification in UI (~50%) - filter; sort
  - Date by far the most common sort attribute (vs. best-match)
    - Importance of time, people, episodes in human memory
    - Few searches for “best match”; many other criteria
  - Need for “abstractions” – date, people, kind
  - Rich client-side interface
    - Support fast iteration/refinement
    - Fast filter-sort-scroll vs. next-next-next

- **Interesting reviews from SIGIR 😊**

- **Practice:** XP and Vista desktop search

*SIGIR 2009*
Re-Finding on the Web

- 50-80% page visits are re-visits
- 30-50% of queries are re-finding queries

Data from Teevan et al., SIGIR 2007

<table>
<thead>
<tr>
<th></th>
<th>Repeat Click</th>
<th>New Click</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeat Query</td>
<td>33%</td>
<td>29%</td>
</tr>
<tr>
<td>New Query</td>
<td>67%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Total = 43%

- Big opportunity to support re-finding on Web
- Models to combine Web rank w/ personal history of interaction
- Interfaces to support finding and re-finding

SIGIR 2009
**Personalization**

- **Today:** People get the same results, independent of current session, previous search history, etc.
- **PSearch** [Teevan et al. 2005]: Uses rich client-side model of a user to personalize search results

**User profile:**

- Content
- Interaction history
Personalization

- Building a User Profile
  - Type of information
    - Content: Past queries, web pages, desktop
    - Behavior: Visited pages, explicit feedback
  - Time frame: Short term, long term
  - Who: Individual, group
  - Where the profile resides:
    - Local: Richer profile, improved privacy [but, increasingly rich public data]
    - Server: Richer communities, portability

- Using the User Profile
  - Ranking
  - Query support
  - Result presentation
Personalization

- Ranking algorithm [Teevan et al. 2007]
  - Linear combination of scores from: content match, history of interaction, Web ranks

- When to personalize [Teevan et al. 2008, in press]
  - Personalization works well for some queries, but not others
  - Models for predicting when to personalize using features of the query and query-user interaction

- Evaluating personalized search [Fox 2005; Teevan et al. in press]
  - What’s relevant for you
    - Explicit judgments (offline and in situ)
    - Implicit “judgments” from behavioral interaction
    - Linking explicit and implicit judgments

Curious Browser Study (~4k)
* 45% w/ just click
* 75% w/ click + dwell + session
Categorization and Metadata

- **Algorithms and applications**
  - Reuters, Web - fast SVM algorithm [Dumais et al. 1998, 2000]
  - Junk email [Sahami et al. 1998]
    - Domain-specific feature engineering
    - Constantly changing content (both ham and spam)

- **Using metadata for ranking** [Bennett et al.]

- **Using metadata in UX**
  - Tight coupling search & browse – e.g., SIS, Phlat [Dumais et al. 2003]
  - Faceted-metadata in many verticals -> Web? [Teevan et al. 2008]
  - Information theoretic models of search/navigation [Downey et al. 2008]

- **Leveraging relations among documents**
Future Challenges

- **Dynamic information environments** [Adar et al., Elsas et al.]
  - Content changes (e.g., news, blogs, lifelogs ... much more general)
  - People re-visit, re-query, re-find
  - IR opportunities ... crawling, doc and user representation, ranking, etc.
  - Interesting historically and socially

- **Data/Evaluation**
  - Data as valuable resource
  - Large-scale log data
  - Operational systems and a “Living Laboratory”
  - IR opportunities ... representations, ranking, etc.

- **Thinking outside the traditional IR boxes**
  - Better understanding of users and application domains
  - Collaborations across disciplinary boundaries
Information Dynamics

Microsoft Research Homepage

1996

2009

SIGIR 2009
Information Dynamics

My Homepage

1998

1998

2008

2008

SIGIR 2009

Susan Dumais

Research Activities:

What's New:

Workshops, Collaborations and Papers:
Information Dynamics

Content Changes

User Visitation/ReVisitation

Today’s Browse and Search Experiences

But, ignores ...
Dynamics and Search

- Improved crawl policy (common use)
- Improved ranking using temporal IR models [Elsas and Dumais]
- Queries have different temporal patterns
- Pages have different rates of change
- Document priors (using temporal vs. link structure)
- Terms have different longevity
- Some are always on the page; some transient
- Show change in snippets
- More general browser support [Teevan et al. 2009]

SIGIR 2009

Welcome – Please join us in Boston | SIGIR 2009

The SIGIR 2009 conference opens in just over a week in Boston, Massachusetts, at the Sheraton Boston Hotel and Northeastern University. The conference is chock full of exciting events and registrations are strong and still growing. We are looking forward to an exciting week.

New content: Please join your colleagues by starting the conference with a free continental breakfast in the Sheraton Hotel, Back Bay A&B, from 7:00am to 8:20am on Monday July 20.
sigir2009.org
Data and Evaluation

- Data as a critical resource

- Shared IR data resources typically consist of:
  - Static collection of documents and queries
  - Judgments of Q-Doc in isolation
  - Judgments with limited context (just the current query)
  - Judges (who are usually not the searcher)

  ... and these resources often shape the questions we ask

- Search is an inherently interactive and iterative process, so user interaction data, is an especially important resource for the IR community
  - Large-scale log data
  - Operational system as an experimental platform
Data and Evaluation

- **Large-scale log data**
  - Understanding how user interact with existing systems
    - What they are trying to do; Where they are failing; etc.
  - Implications for: models, and interactive systems
  - Lemur Query Log Toolbar – developing a community resource!

- **Operational systems as an experimental platform**
  - Can also conduct controlled experiments *in situ*
    - Interleave results from different methods [Radlinski & Joachims 2005]
    - A/B testing -- Data vs. the “hippo” [Kohavi 2008]
  - Important in: linking offline and interactive results, understanding effect sizes, relations among results (and other page components), etc.
  - Can we build such a “Living Laboratory”?

- Replicability in the face of changing content, users, queries

SIGIR 2009
Opportunities

- Continued improvements in representation and ranking
- Think outside the traditional IR boxes !!!
  - Develop a better understanding of users, and their tasks
  - Design and evaluate interactive systems to support this
- Importance of
  - New data resources
  - Interdisciplinary perspective

SIGIR 2009
Thanks (again)!

Bell Labs

MSR, CLUES (Context, Learning and User Experience In Search)

SIGIR 2009