Using Context to Support Searchers in Searching

Susan Dumais

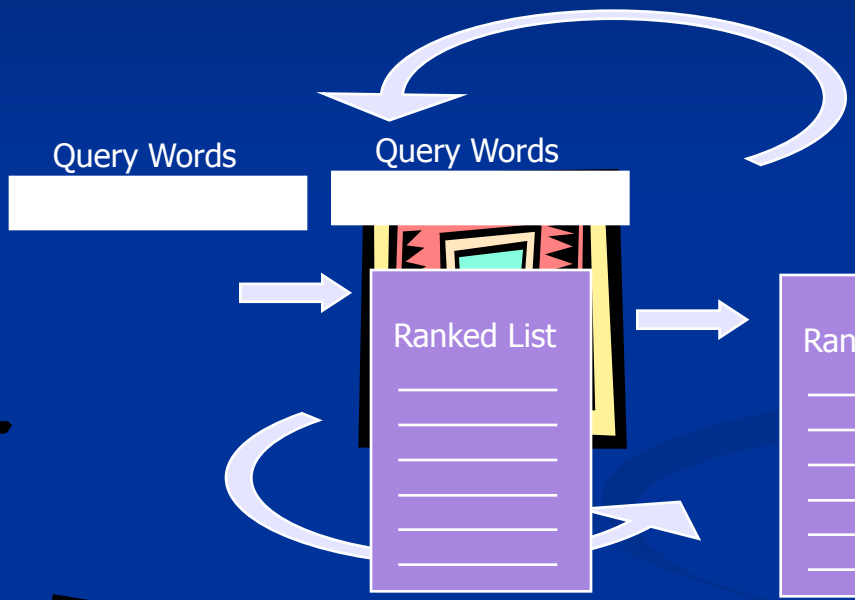
Microsoft Research

http://research.microsoft.com/~sdumais

ACL/HLT – June 18, 2008

Using Context to Support Searchers

User Context



Document Context

Task/Use Context

Web Info through the Years

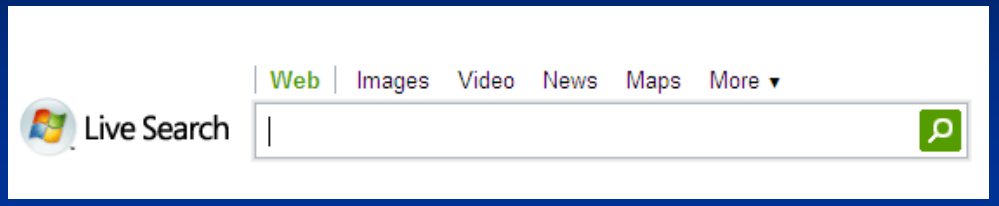
What's available

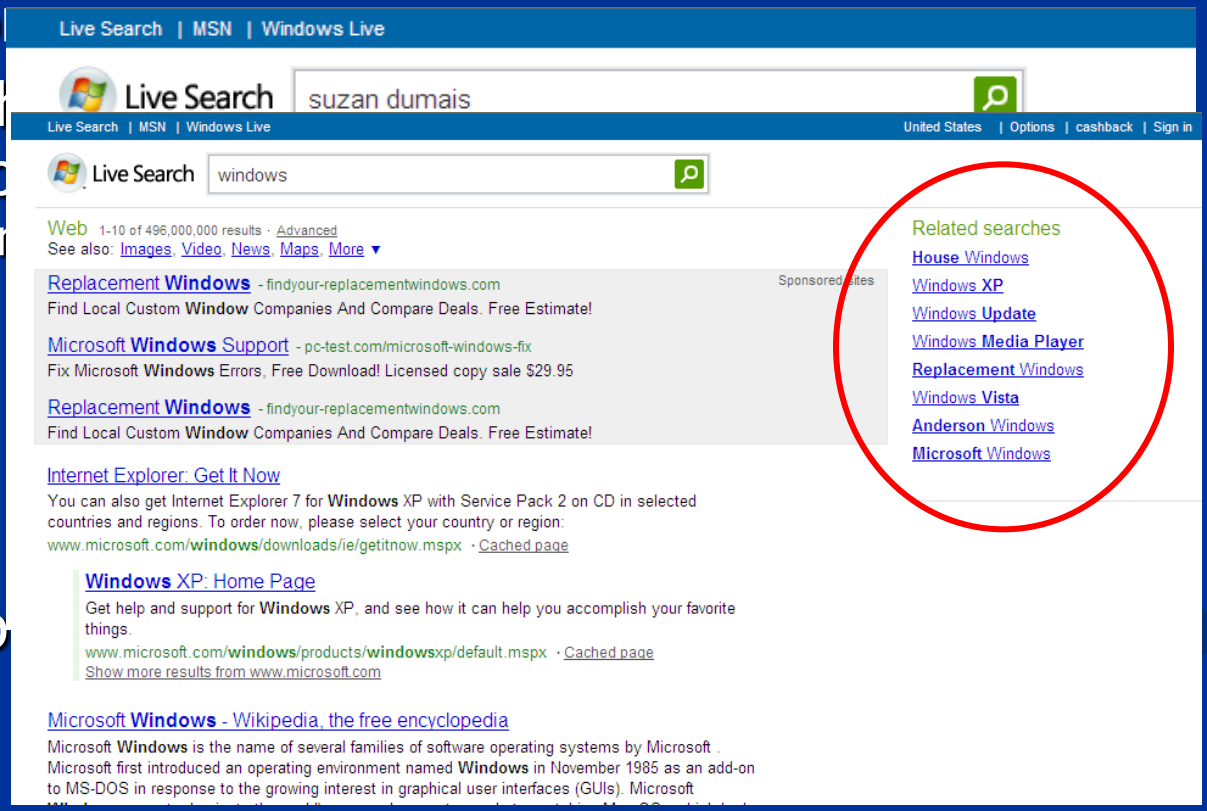
- Number of pages indexed
  - 7/94 Lycos –
  - 95 – 10^6 millions
  - 97 – 10^7
  - 98 – 10^8
  - 01 – 10^9 billions
  - 05 – 10^10 …
- Types of content
  - Web pages, newsgroups
  - Images, videos, maps
  - News, blogs, spaces
  - Shopping, local, desktop
  - Books, papers
  - Health, finance, travel …

How it's accessed

Some Support for Searchers

- The search box
- Spelling suggestions
- Query suggestio
- Advanced search operators and op (e.g., “”, +/-, site:, lan filetype:, intitle:)
- Richer snippets
- But, we can do b 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)

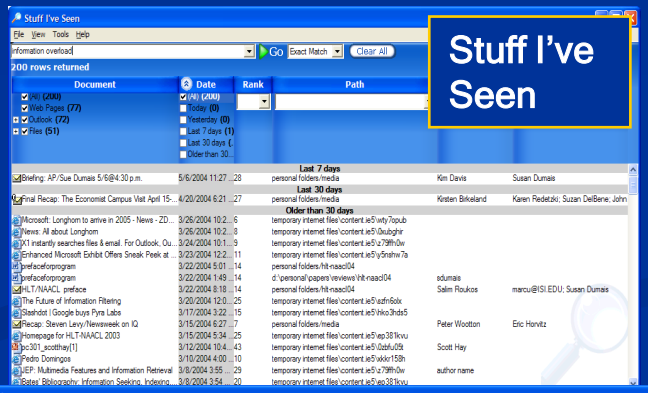
Dumais et al., SIGIR 2003

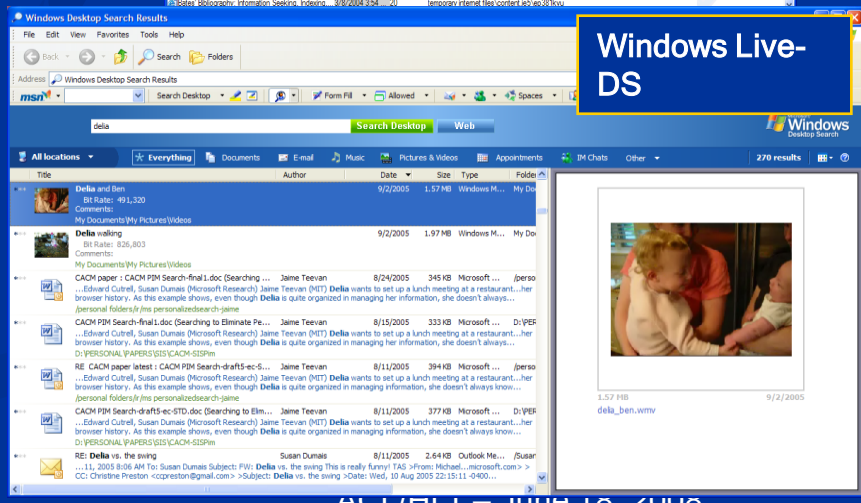
SIS: Stuff I've Seen

- 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

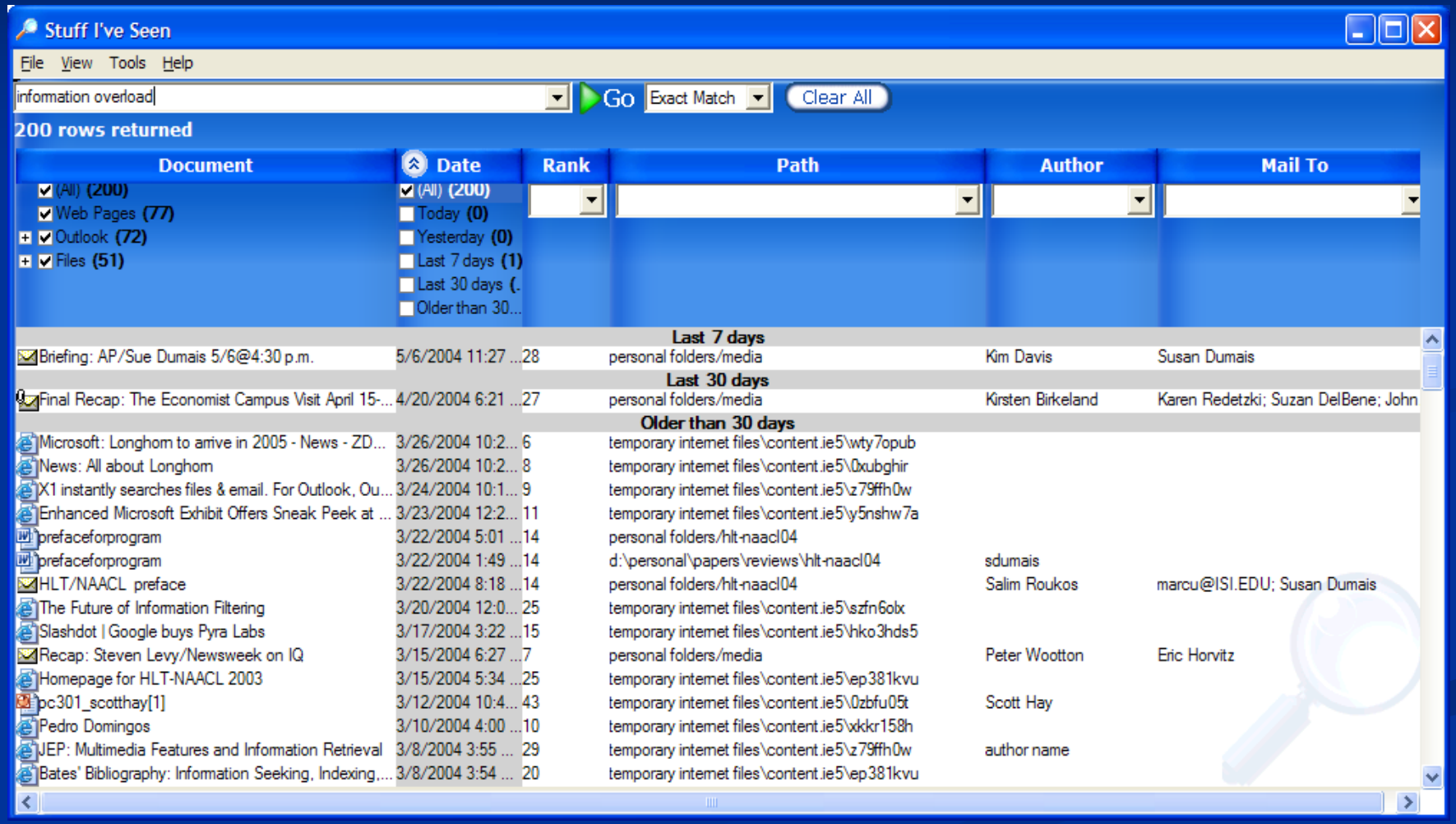
➡ *Vista Desktop Search (and Live Toolbar)*

Also, Spotlight, GDS, X1, …





SIS Demo


Stuff I've Seen
File View Tools Help
information overload
Go
Exact Match
Clear All
200 rows returned
Document
Date
Rank
Path
Author
Mail To
(All) (200)
Web Pages (77)
Outlook (72)
Files (51)
(All) (200)
Today (0)
Yesterday (0)
Last 7 days (1)
Last 30 days (.
Older than 30...
Last 7 days
Briefing: AP/Sue Dumais 5/6@4:30 p.m. 5/6/2004 11:27 ... 28 personal folders/media Kim Davis Susan Dumais
Last 30 days
Final Recap: The Economist Campus Visit April 15-... 4/20/2004 6:21 ... 27 personal folders/media Kirsten Birkeland Karen Redetzki; Suzan DelBene; John
Older than 30 days
Microsoft: Longhorn to arrive in 2005 - News - ZD... 3/26/2004 10:2... 6 temporary internet files\content.ie5\wty7opub
News: All about Longhorn 3/26/2004 10:2... 8 temporary internet files\content.ie5\0xubghir
X1 instantly searches files & email. For Outlook, Ou... 3/24/2004 10:1... 9 temporary internet files\content.ie5\z79ffh0w
Enhanced Microsoft Exhibit Offers Sneak Peek at ... 3/23/2004 12:2... 11 temporary internet files\content.ie5\y5nshw7a
prefaceforprogram 3/22/2004 5:01 ... 14 personal folders/hlt-naacl04
prefaceforprogram 3/22/2004 1:49 ... 14 d:\personal\papers\reviews\hlt-naacl04 sdumais
HLT/NAACL preface 3/22/2004 8:18 ... 14 personal folders/hlt-naacl04 Salim Roukos marcu@ISI.EDU; Susan Dumais
The Future of Information Filtering 3/20/2004 12:0... 25 temporary internet files\content.ie5\szfn6olx
Slashdot | Google buys Pyra Labs 3/17/2004 3:22 ... 15 temporary internet files\content.ie5\hko3hds5
Recap: Steven Levy/Newsweek on IQ 3/15/2004 6:27 ... 7 personal folders/media Peter Wootton Eric Horvitz
Homepage for HLT-NAACL 2003 3/15/2004 5:34 ... 25 temporary internet files\content.ie5\ep381kvu
pc301_scotthay[1] 3/12/2004 10:4... 43 temporary internet files\content.ie5\0zbfu05t Scott Hay
Pedro Domingos 3/10/2004 4:00 ... 10 temporary internet files\content.ie5\xkkr158h
JEP: Multimedia Features and Information Retrieval 3/8/2004 3:55 ... 29 temporary internet files\content.ie5\z79ffh0w author name
Bates' Bibliography: Information Seeking, Indexing,... 3/8/2004 3:54 ... 20 temporary internet files\content.ie5\ep381kvu

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

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

Query characteristics

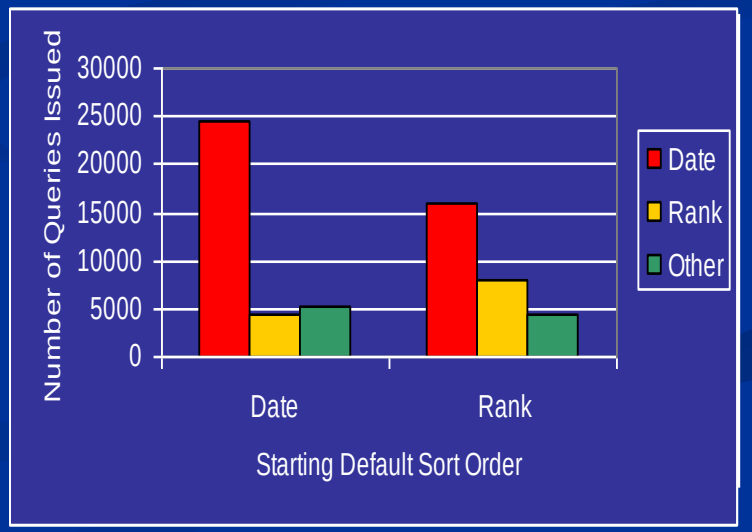
- Short queries (1.6 words)
- Few advanced operators or fielded search in query box (~7%)
- Many advanced operators and query iteration in UI (48%)
  - Filters (type, date); modify query; re-sort results

SIS Usage Data, cont'd

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

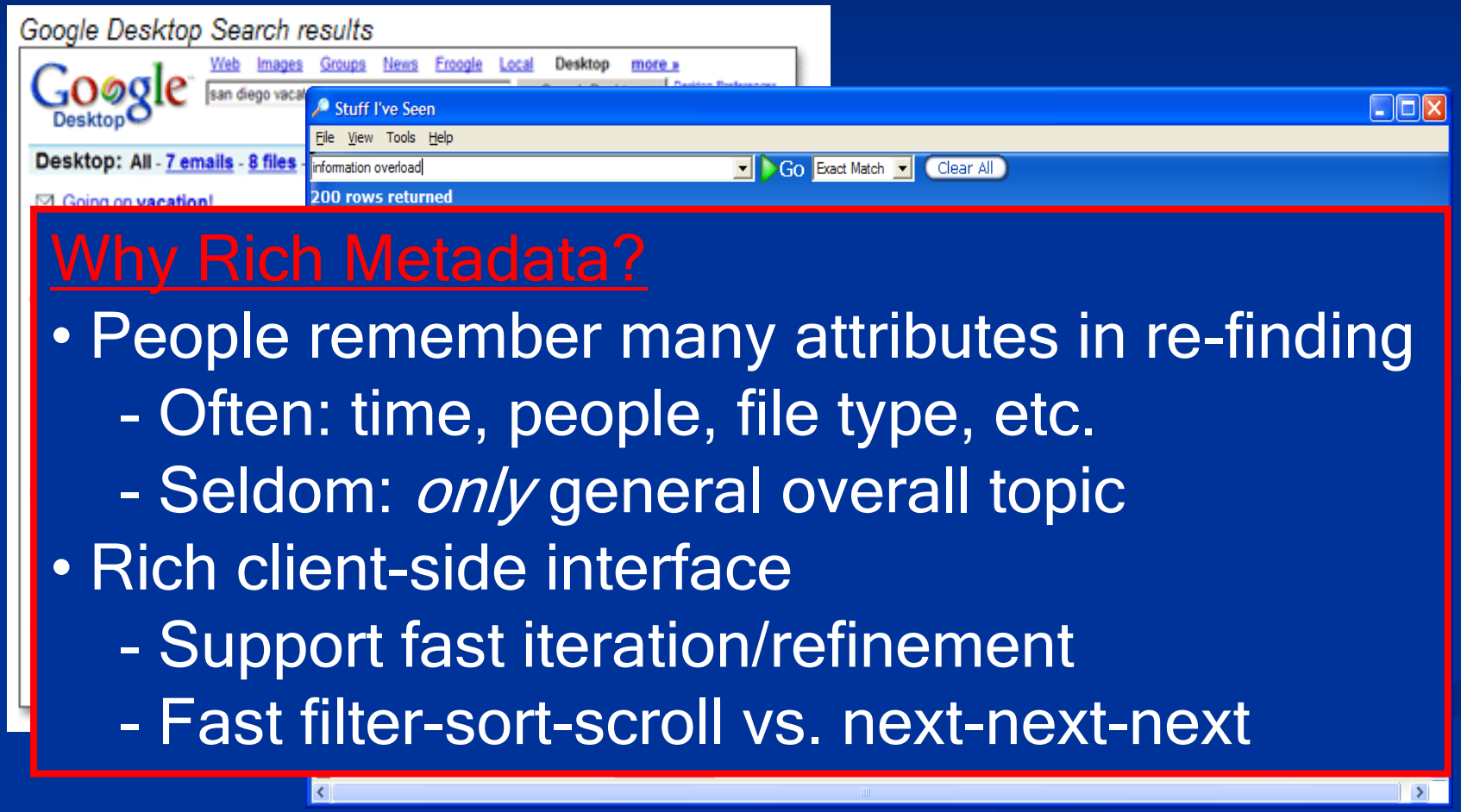


SIS Usage Data, cont'd

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

Teevan et al., SIGIR 2007

Re-finding on the Web

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

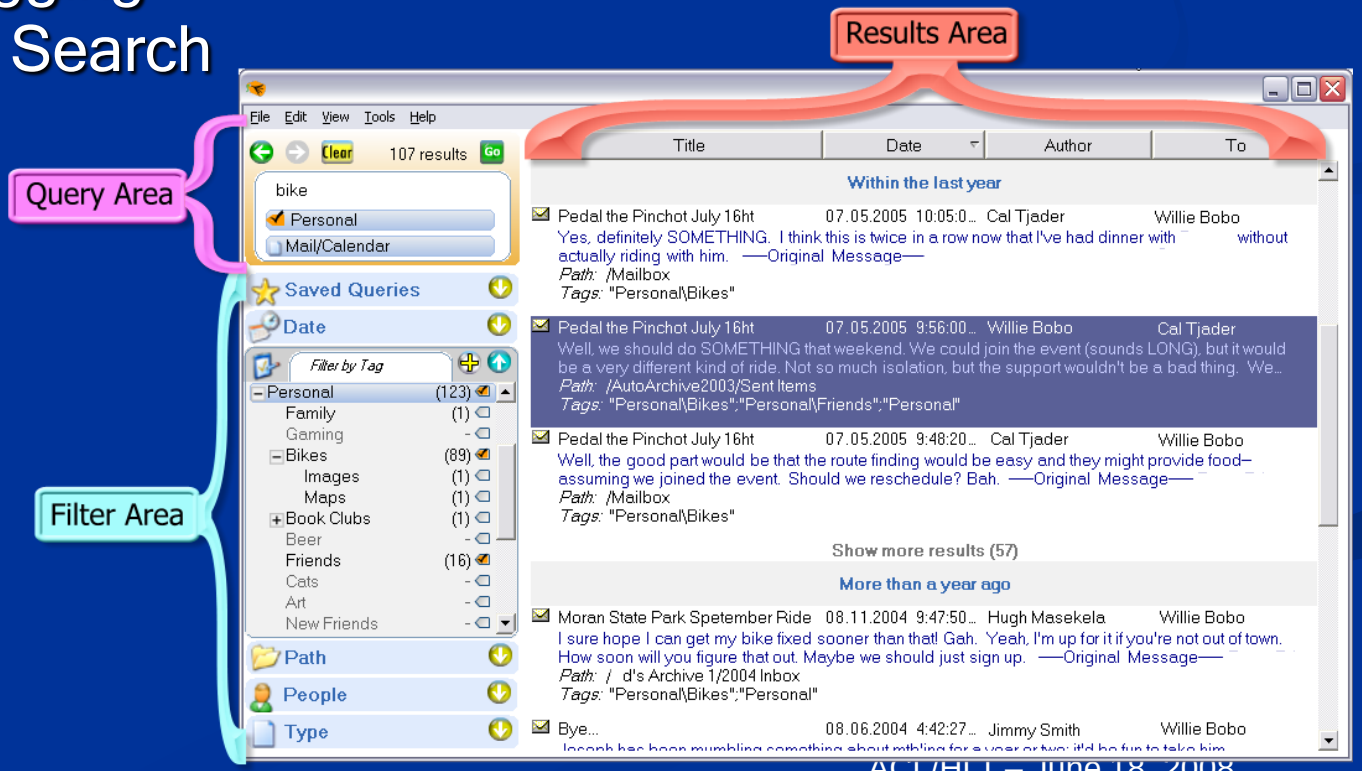
Table 1. A classification of different query types.

All queries: 13,060 queries (100%)	Overlapping Click Queries – 5072 queries (39%)			
	Equal Click Queries – 3777 (29%)		Some Common Clicks 1295 (10%)	No Common Clicks 7988 (61%)
	Single Identical Click 3737 (29%)	Multiple Identical Clicks 40 (< 1%)		
Equal Query Queries 4256 (33%)	Navigational Queries 3100 (24%)	36 (< 1%)	635 (5%)	485 (4%)
Different Query 8804 (67%)	637 (5%)	4 (< 1%)	660 (5%)	7503 (57%)

Cutrell et al., CHI 2006

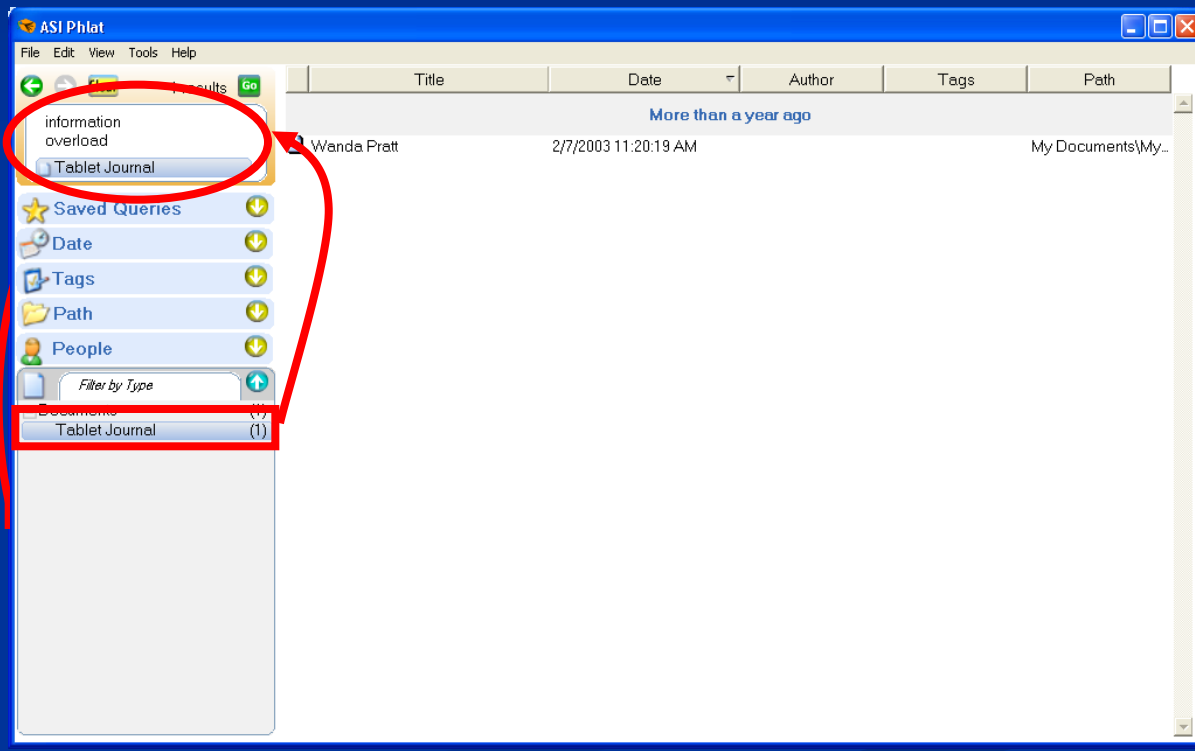
Phlat: Search and Metadata

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



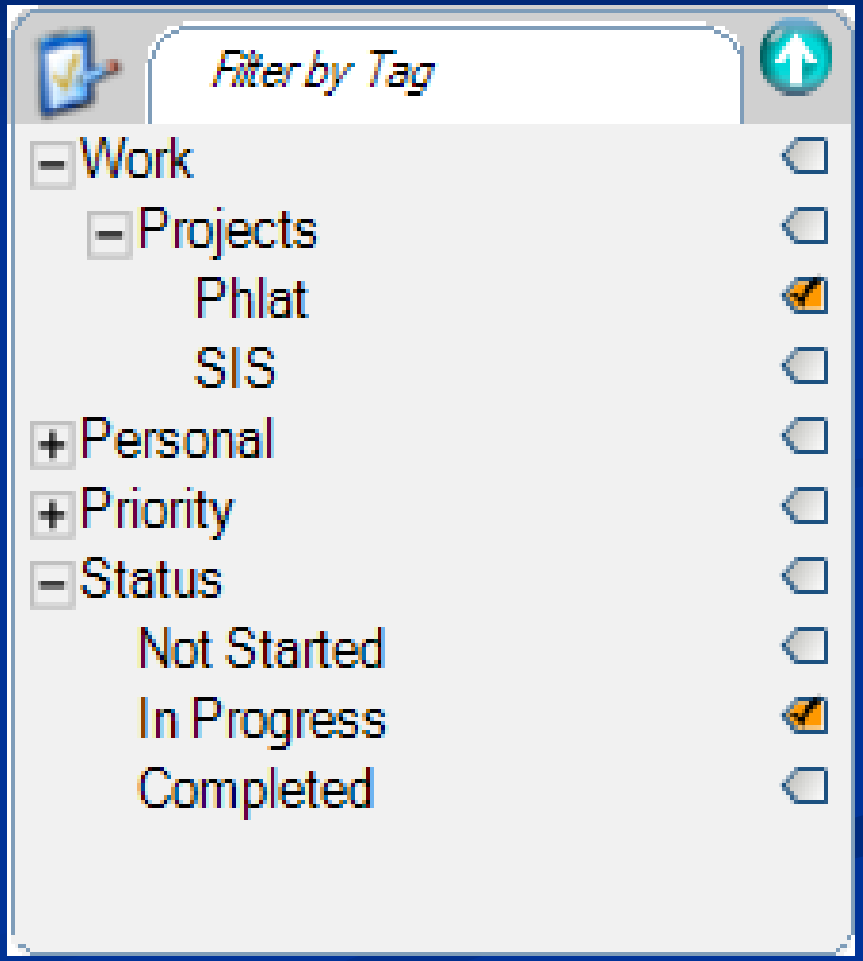
Phlat: Faceted metadata

- 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



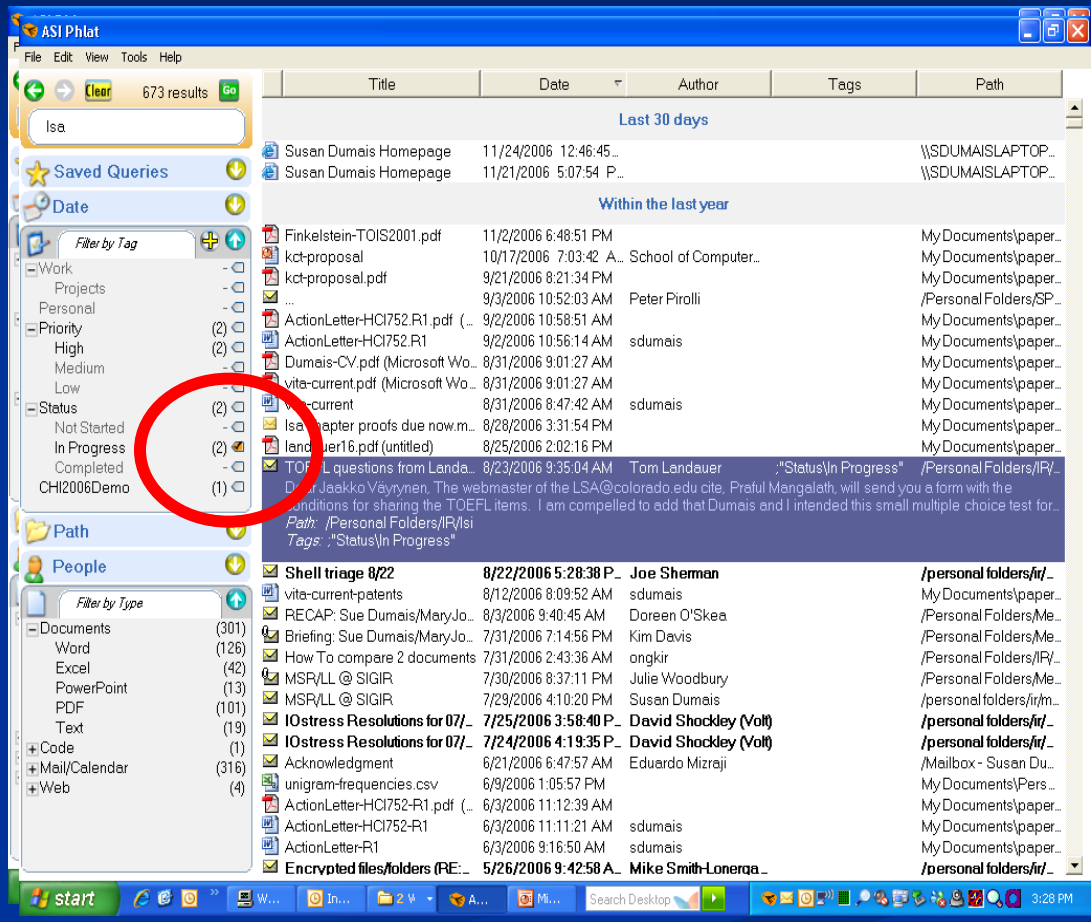
Phlat: Tagging

- 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

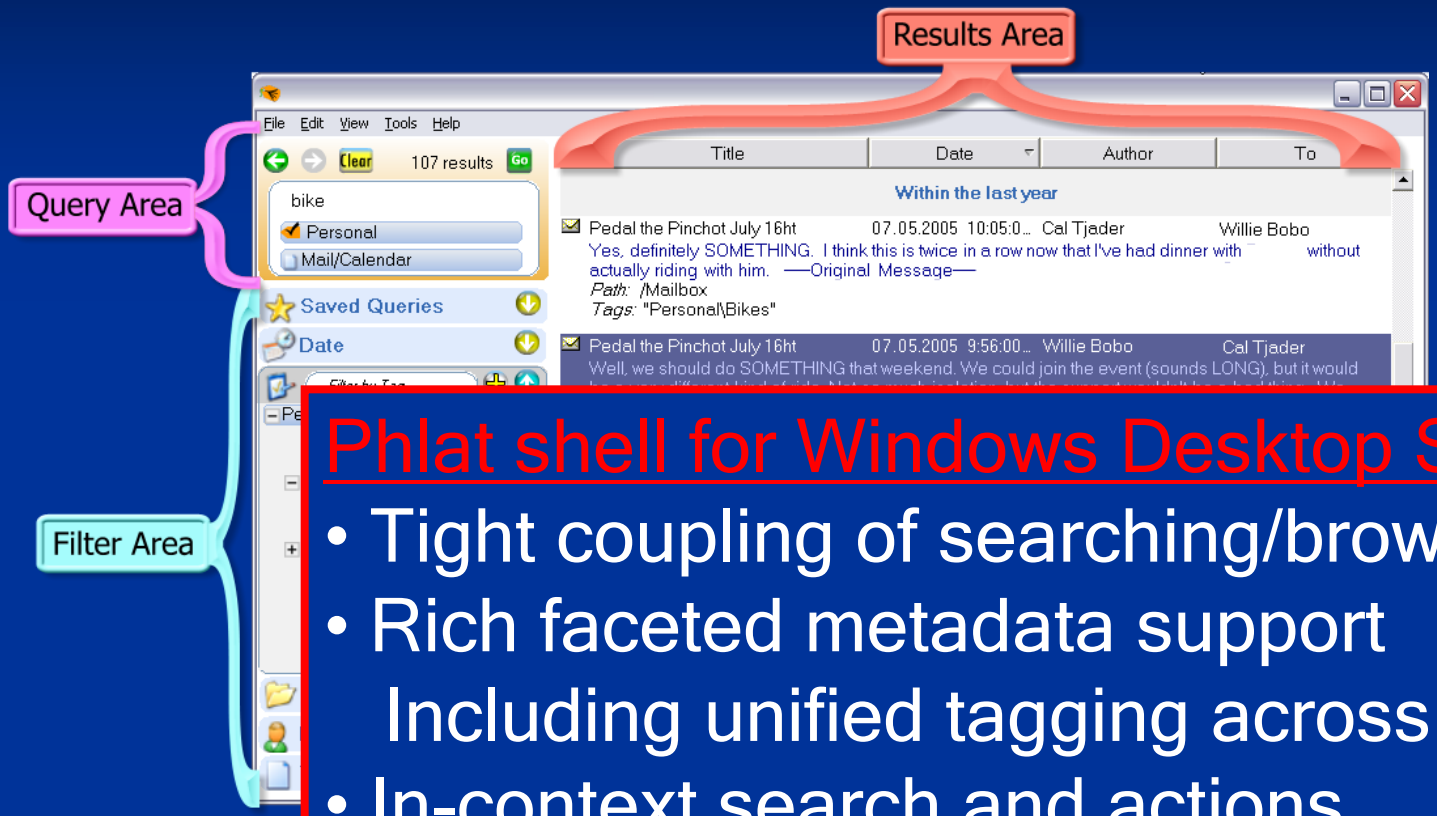


Phat: In-Context Search

- 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



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

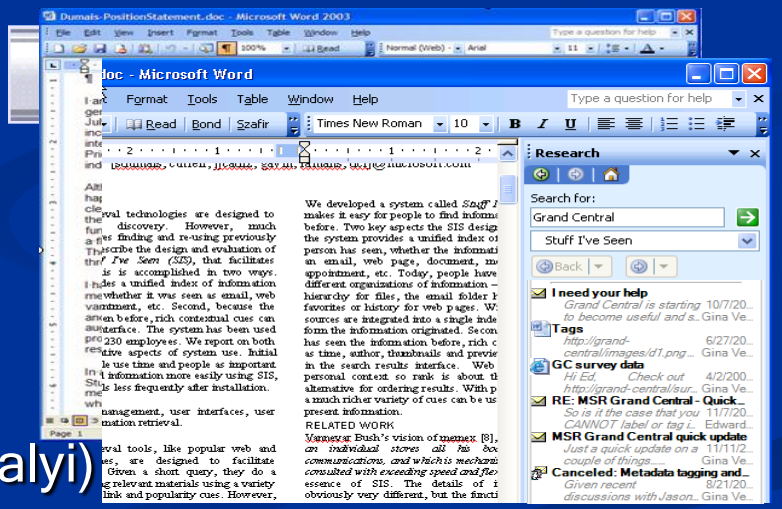
Download: http://research.microsoft.com/adapt/phlat

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

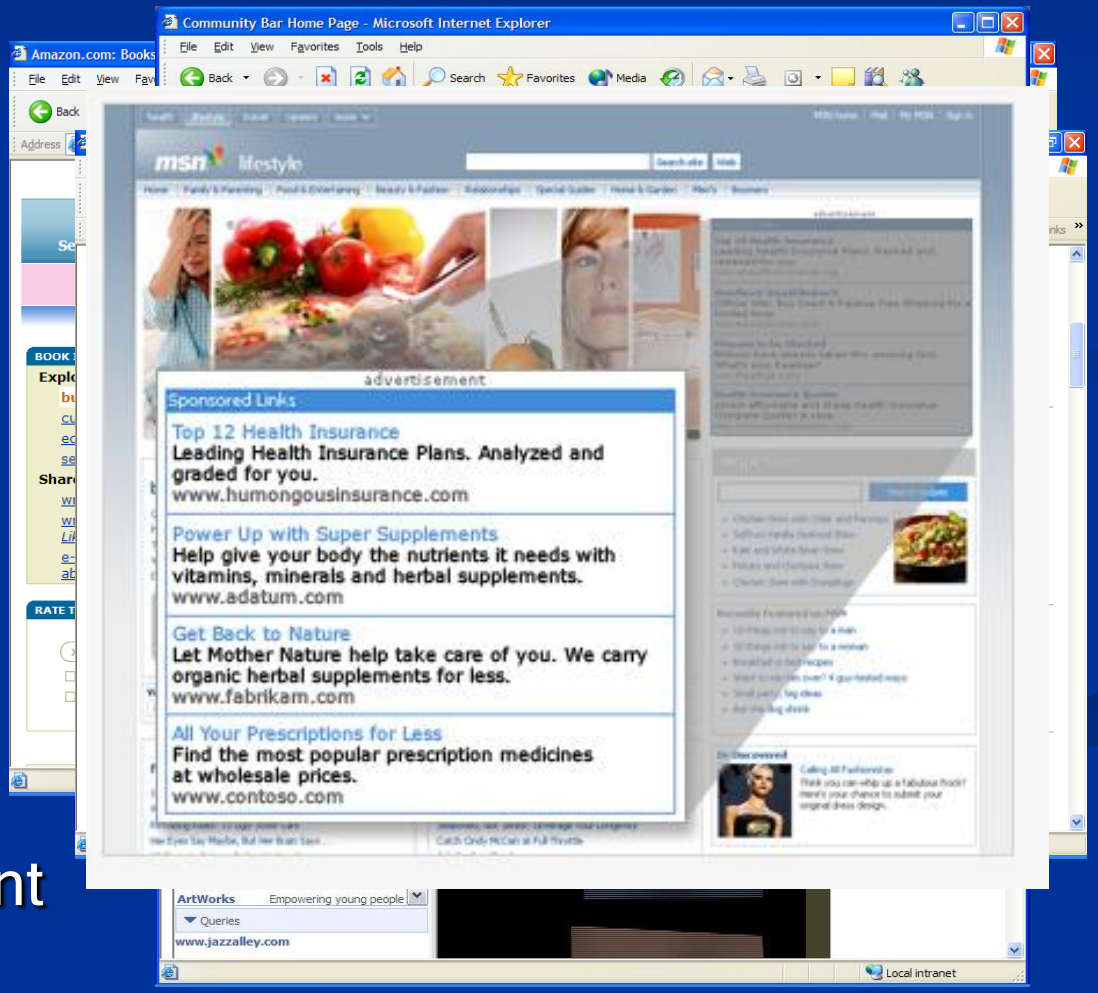
- Search is not the end goal …
- Support information access in the context of ongoing activities (e.g., writing talk, finding out about, planning trip, buying, monitoring, etc.)
  - Search always available
  - Search from within apps (keywords, regions, full doc)
  - Show results within app
  - Maintains "flow" (Csikszentmihalyi)
  - Can improve relevance



Documents as (a simple) Context

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

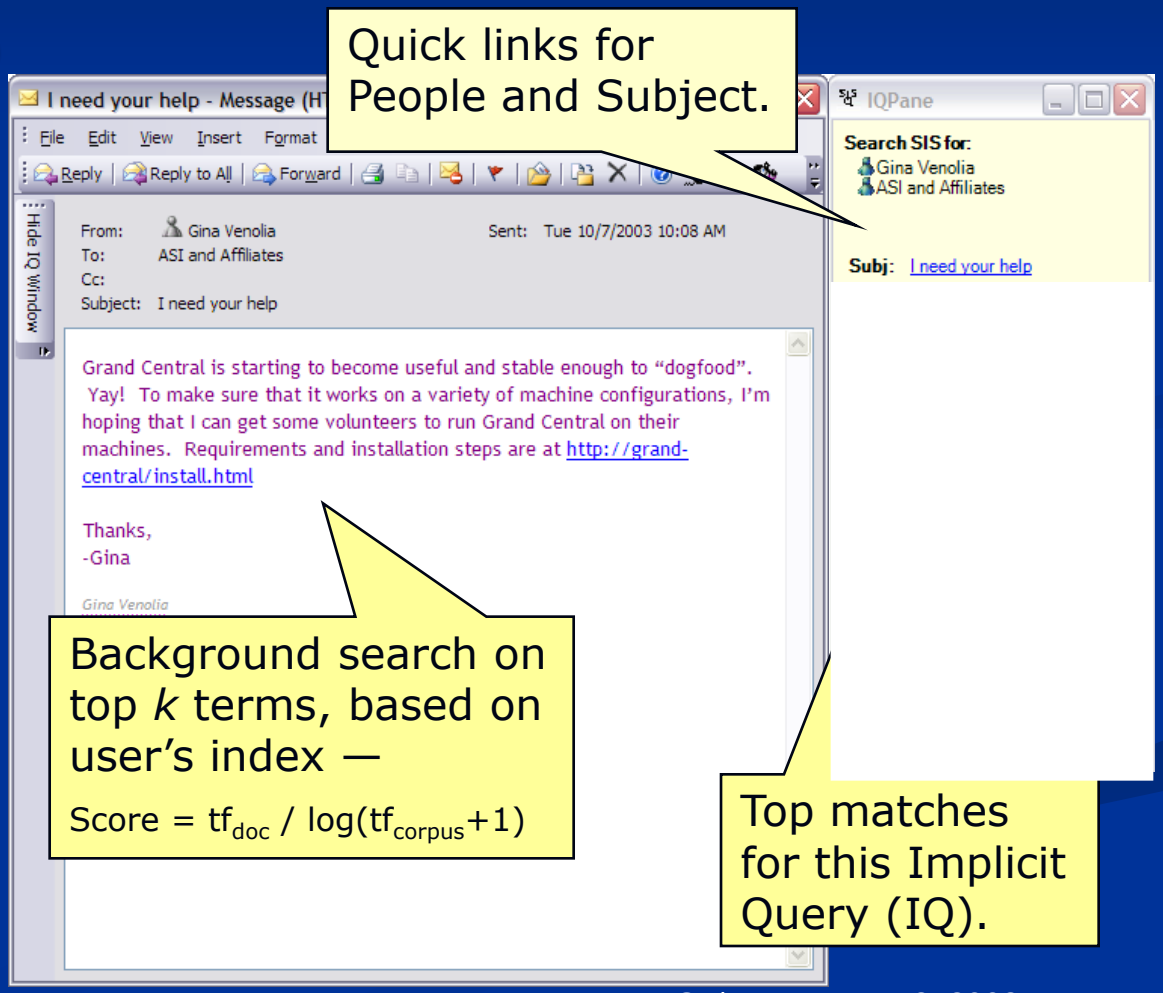
- Recommendations
  - People who bought this also bought …
- Contextual Ads
  - Ads relevant to page
- Community Bar
  - Notes, Chat, Tags, Inlinks, Queries
- Implict Queries (IQ)
  - Also Y!Q, Watson, Rememberance Agent



Dumais et al., SIGIR 2004

Document Contexts (Implicit Query, IQ)

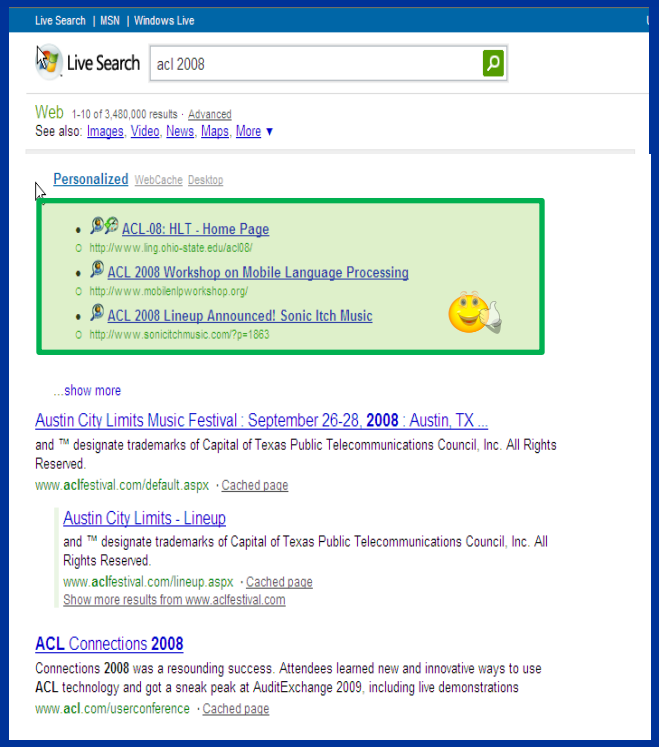
- Proactively find info related to item being read/created
  - Quick links
  - Related content
- Challenges
  - Relevance, fine
  - When to show? (useful)
  - How to show? (peripheral awareness)



Teevan et al., SIGIR 2005

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



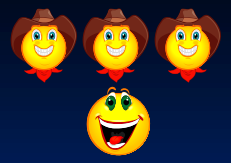
- Building a user profile

- 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

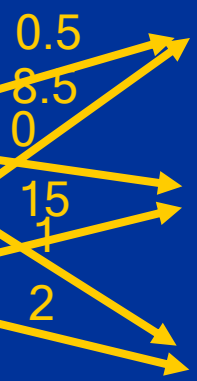


Building a User Profile

PSearch

- 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

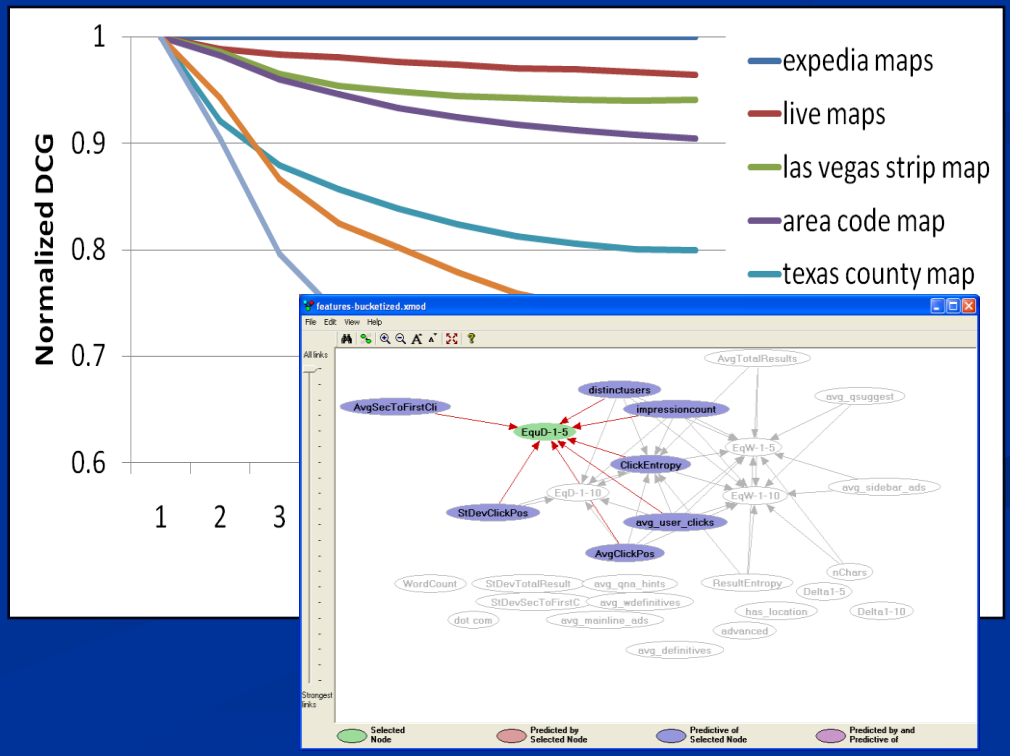
Personalized Ranking



- Personal Rank = *f(Cont, Beh, Web)*
  - Pers_Content Match: *sim(result, user_content_profile)*
  - Pers_Behavior Match: *visited URLs*
  - Web Match: *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)*"

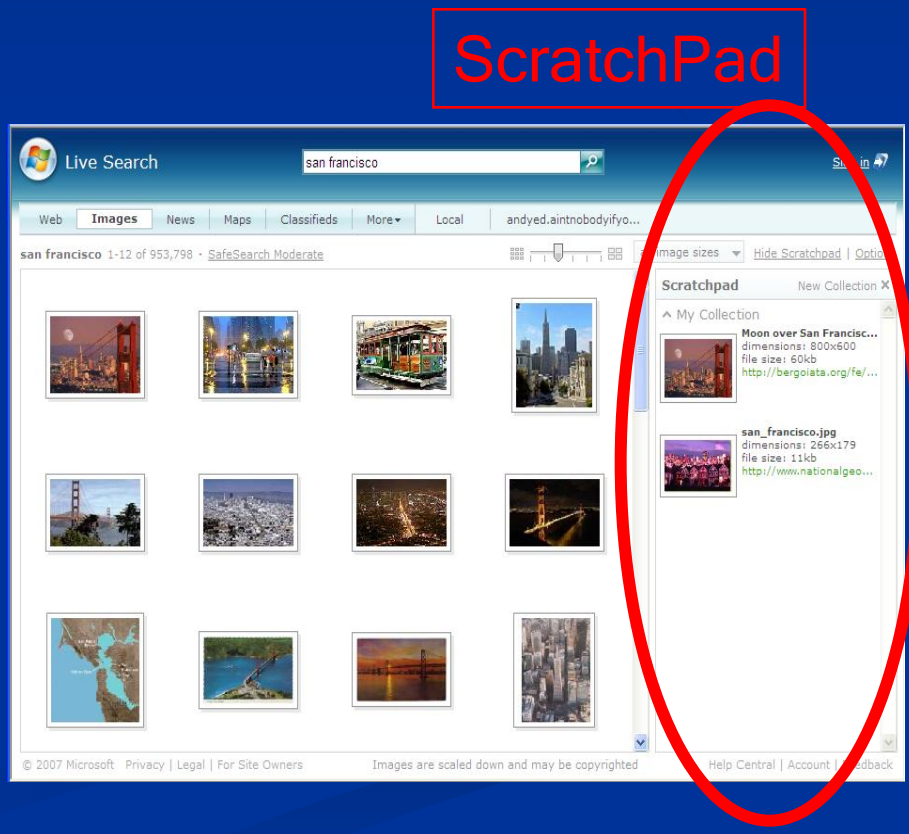


More Personalized Search

- 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



Morris et al., UIST 2007

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, ever
    understanding medical info; researchi
    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)

SearchTogether

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

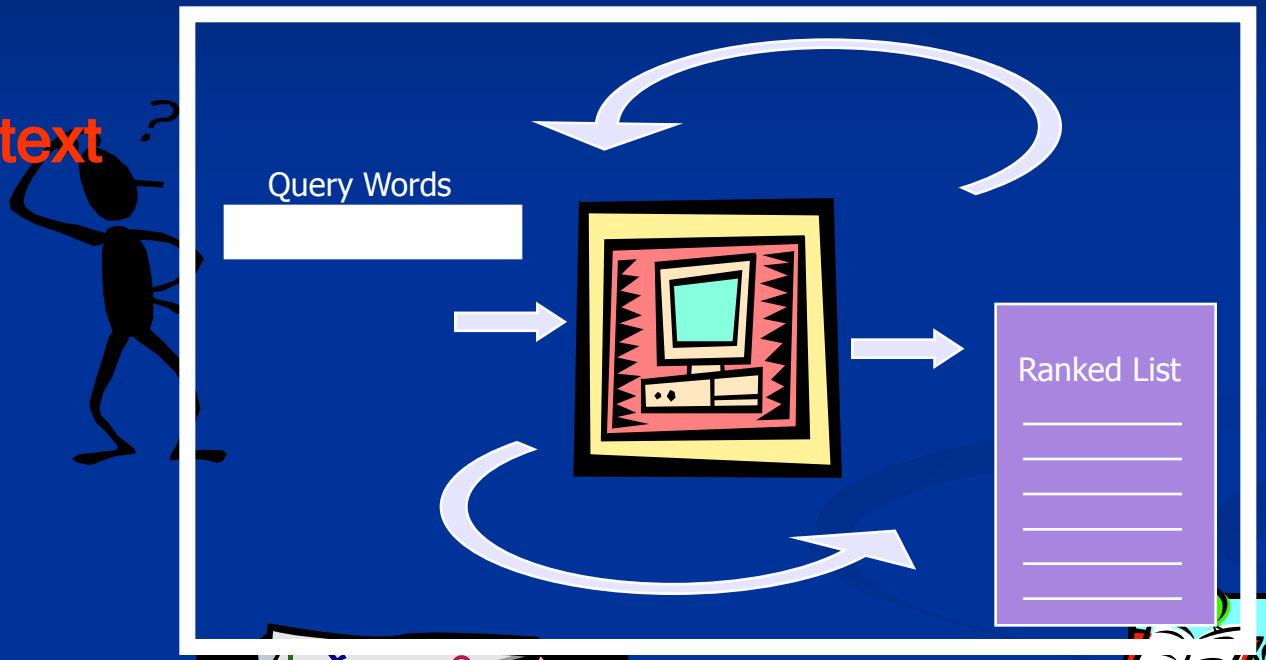
Looking Ahead …

- 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

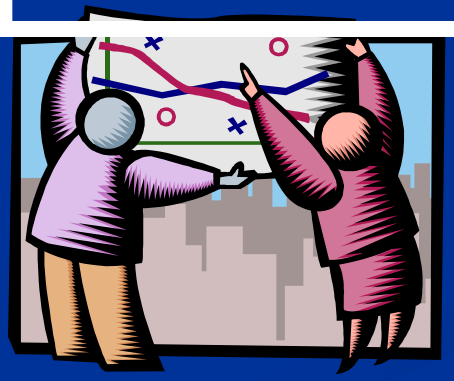
Think Outside the IR Box(es)

User Context



Document Context

Task/Use Context



Thank You !

- Questions/Comments …
- More info, http://research.microsoft.com/~sdumais
- Windows Live Desktop Search, http://toolbar.live.com
- Phlat, http://research.microsoft.com/adapt/phlat
- Search Together, http://research.microsoft.com/searchtogether/

References

- Stuff I've Seen
  - S. T. Dumais, E. Cutrell, J. J. Cadiz, G. Jancke, R. Sarin & D. C. Robbins (2003). Stuff I've Seen: A system for personal information retrieval and re-use. *SIGIR 2003*.
  - *Download: http://toolbar.live.com and Vista Search*
- Phlat
  - E. Cutrell, D. C. Robbins, S. T. Dumais & R. Sarin (2006). Fast, flexible filtering with Phlat - Personal search and organization made easy. *CHI 2006.*
  - *Download: http://research.microsoft.com/adapt/phlat*
- Memory Landmarks
  - M. Ringel, E. Cutrell, S. T. Dumais & E. Horvitz (2003). Milestones in time: The value of landmarks in retrieving information from personal stores. *Interact 2003*.
- Personalized Search
  - J. Teevan, S. T. Dumais & E. Horvitz (2005). Personalizing search via automated analysis of interests and activities. *SIGIR 2005*.
- Implicit Queries
  - S. T. Dumais, E. Cutrell, R. Sarin & E. Horvitz (2004). Implicit queries (IQ) for contextualized search. *SIGIR 2004.*
- Revisitation on Web
  - J. Teevan, E. Adar, R. Jones & M. Potts (2007). Information re-retrieval. *SIGIR 2007.*
- InkSeine
  - K. Hinckley, S. Zhao, R. Sarin, P Baudisch, E. Cutrell & M. Shilman (2007). InkSeine: In situ search for active note taking. *CHI 2007.*
  - *Download: http://research.microsoft.com/inkseine/*
- Search Together
  - M. Morris & E. Horvitz (2007). Search Together: An interface for collaborative web search. *UIST 2007.*
  - *Download: http://research.microsoft.com/searchtogether/*