

# Thinking Outside the (Search) Box

*Susan Dumais*

Microsoft Research

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

# Web Info through the Years

## What's available

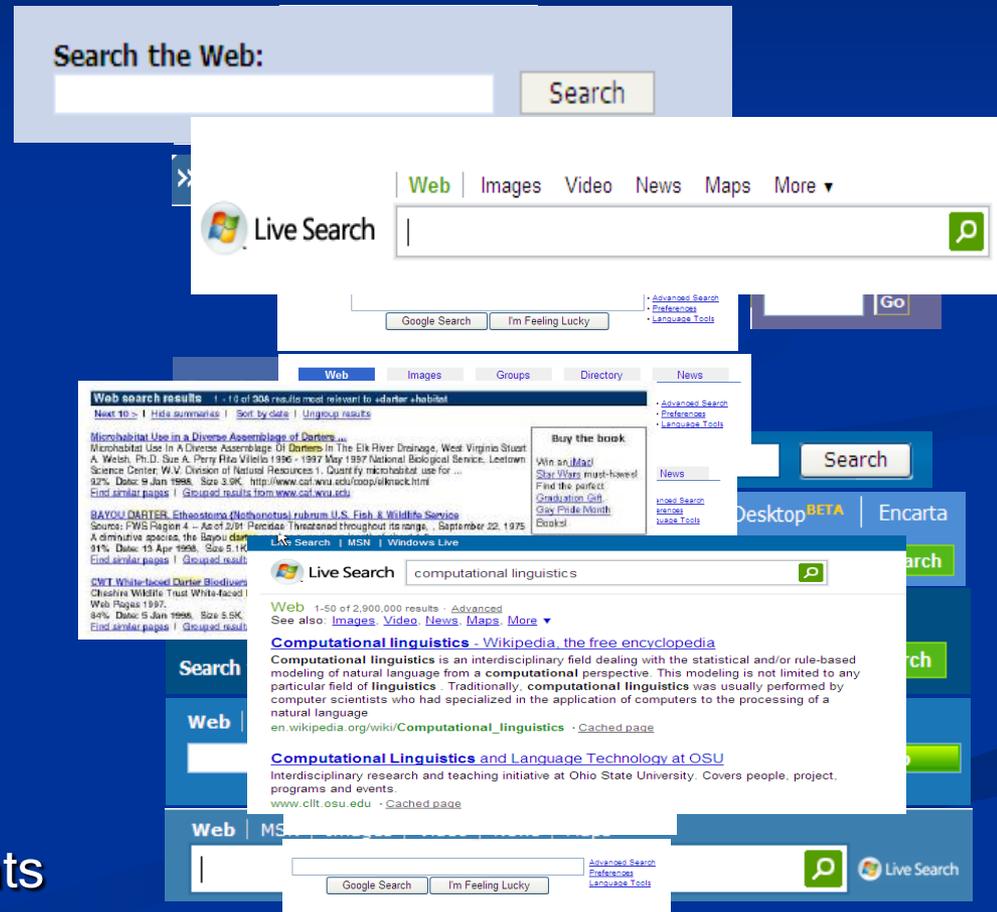
### ■ Number of pages indexed

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

### ■ Types of content

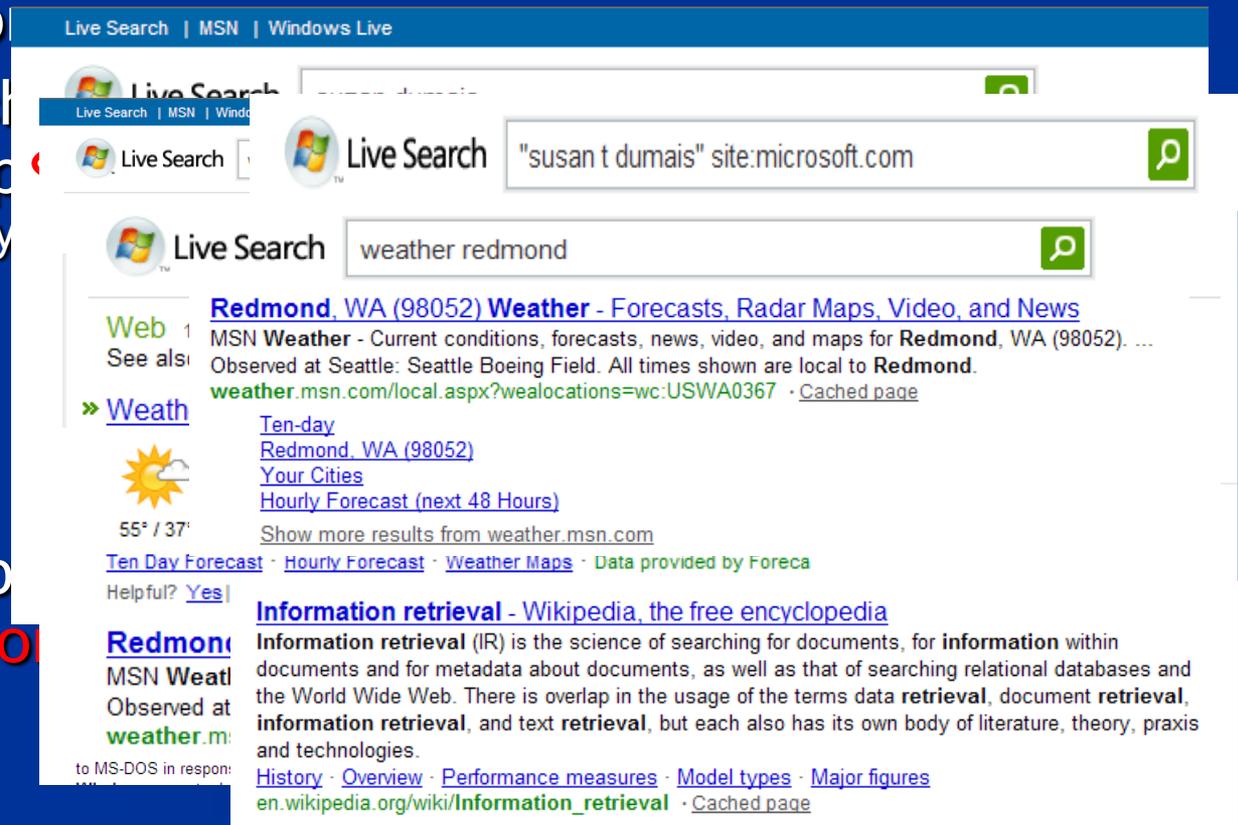
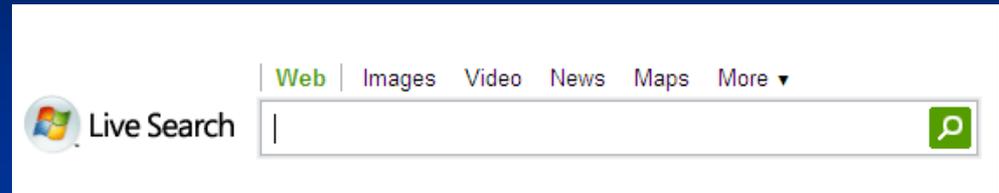
- Web pages, newsgroups
- Images, videos, maps
- News, blogs, spaces
- Shopping, local, desktop
- Books, papers, many formats
- Health, finance, travel ...

## How it's accessed



# Supporting Searchers

- The search box
- Spelling suggestions
- Query suggestions
- Advanced search operators and options (e.g., “”, +/-, site:, filetype:)
- Inline answers
- Richer snippets
- But, we can do better: **understanding context**



# Search and Context

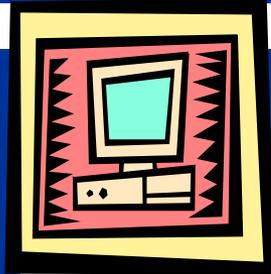
User Context



Query Words



Query Words



Ranked List



Document Context



Task/Use Context



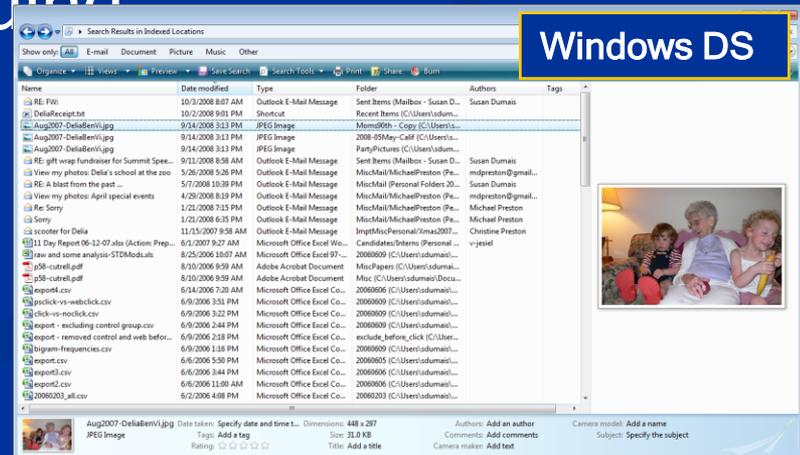
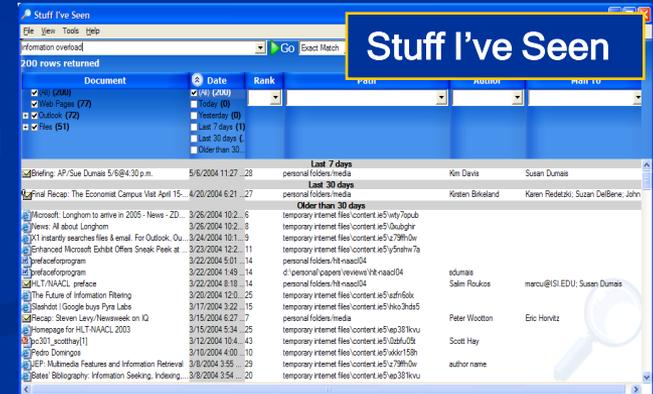
# Search and Context

Research prototypes: extend search algorithmic, capabilities, and user experiences

- User Contexts:
  - Finding and Re-Finding (Stuff I've Seen)
  - Novelty in News (NewsJunkie)
  - Personalized Search (PSearch)
- Document/Domain Contexts:
  - Metadata and search (SIS, Phlat)
  - Visualizing patterns in results (MemoryLandmarks, GridViz)
  - Dynamic information environments (DiffIE)
- Task/Use Contexts:
  - Pages as context (Community Bar, IQ)
  - Richer collections as context (NewsJunkie, PSearch)
  - Understanding, sharing (SearchTogether, InkSeine)

# Stuff I've Seen (SIS)

- Unified index of *stuff you've seen*
  - 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
  - Supports re-finding vs. finding



➔ *Vista Desktop Search (and XP, 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
<input checked="" type="checkbox"/> (All) (200) <input checked="" type="checkbox"/> Web Pages (77) <input checked="" type="checkbox"/> Outlook (72) <input checked="" type="checkbox"/> Files (51)					
<input checked="" type="checkbox"/> (All) (200) <input checked="" type="checkbox"/> Today (0) <input type="checkbox"/> Yesterday (0) <input type="checkbox"/> Last 7 days (1) <input type="checkbox"/> Last 30 days (. <input type="checkbox"/> Older than 30...					
<b>Last 7 days</b>					
<input checked="" type="checkbox"/> Briefing: AP/Sue Dumais 5/6@4:30 p.m.	5/6/2004 11:27 ...28		personal folders/media	Kim Davis	Susan Dumais
<b>Last 30 days</b>					
<input checked="" type="checkbox"/> Final Recap: The Economist Campus Visit April 15-...	4/20/2004 6:21 ...27		personal folders/media	Kirsten Birkeland	Karen Redetzki; Suzan DelBene; John
<b>Older than 30 days</b>					
<input type="checkbox"/> Microsoft: Longhom to arrive in 2005 - News - ZD...	3/26/2004 10:2... 6		temporary internet files\content.ie5\wty7opub		
<input type="checkbox"/> News: All about Longhom	3/26/2004 10:2... 8		temporary internet files\content.ie5\0xubghir		
<input type="checkbox"/> X1 instantly searches files & email. For Outlook, Ou...	3/24/2004 10:1... 9		temporary internet files\content.ie5\z79ffh0w		
<input type="checkbox"/> Enhanced Microsoft Exhibit Offers Sneak Peek at ...	3/23/2004 12:2... 11		temporary internet files\content.ie5\y5nshw7a		
<input type="checkbox"/> prefaceforprogram	3/22/2004 5:01 ...14		personal folders/hlt-naacl04		
<input type="checkbox"/> prefaceforprogram	3/22/2004 1:49 ...14		d:\personal\papers\reviews\hlt-naacl04	sdumais	
<input checked="" type="checkbox"/> HLT/NAACL preface	3/22/2004 8:18 ...14		personal folders/hlt-naacl04	Salim Roukos	marcu@ISI.EDU; Susan Dumais
<input type="checkbox"/> The Future of Information Filtering	3/20/2004 12:0... 25		temporary internet files\content.ie5\szfn6ox		
<input type="checkbox"/> Slashdot   Google buys Pyra Labs	3/17/2004 3:22 ...15		temporary internet files\content.ie5\hko3hds5		
<input checked="" type="checkbox"/> Recap: Steven Levy/Newsweek on IQ	3/15/2004 6:27 ...7		personal folders/media	Peter Wootton	Eric Horvitz
<input type="checkbox"/> Homepage for HLT-NAACL 2003	3/15/2004 5:34 ...25		temporary internet files\content.ie5\vp381kvu		
<input type="checkbox"/> pc301_scotthay[1]	3/12/2004 10:4... 43		temporary internet files\content.ie5\0zbfu05t	Scott Hay	
<input type="checkbox"/> Pedro Domingos	3/10/2004 4:00 ...10		temporary internet files\content.ie5\dkkr158h		
<input type="checkbox"/> JEP: Multimedia Features and Information Retrieval	3/8/2004 3:55 ... 29		temporary internet files\content.ie5\z79ffh0w	author name	
<input type="checkbox"/> Bates' Bibliography: Information Seeking, Indexing,...	3/8/2004 3:54 ... 20		temporary internet files\content.ie5\vp381kvu		

# 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, people); 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
<b>Total</b>	<b>91k items</b>	<b>25.4 Gb</b>
<b>Index</b>		<b>190 Mb</b>
		<b>+1.5 Mb/week</b>

# SIS Usage Data, cont'd

## Characteristics of items opened

- File types opened
  - 76% Email
  - 14% Web pages
  - 10% Files
- Age of items opened
  - 5% today
  - 21% within the last week
  - 47% within the last month
  - 50% of the cases -> 36 days
    - Web: 11 days
    - Mail: 36 days
    - Files: 55 days

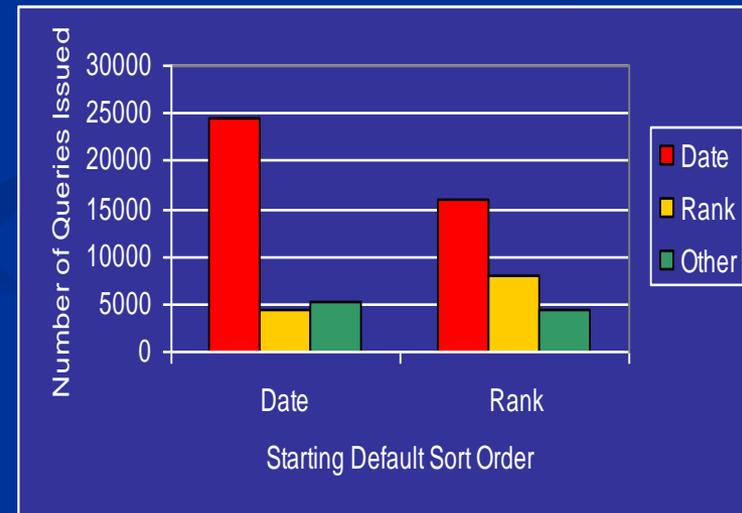
$$\text{Log(Freq)} = -0.68 * \text{log(DaysSinceSeen)} + 2.02$$



# SIS Usage Data, cont'd

## UI Usage

- Small effects of: Top/Side, Previews/NoPreviews
- Large effect of Sort Order:
  - **Date** by far the most common sort field, even for people who had best-match Rank as default
  - Importance of time
  - Few searches for “best” match; many other criteria ...

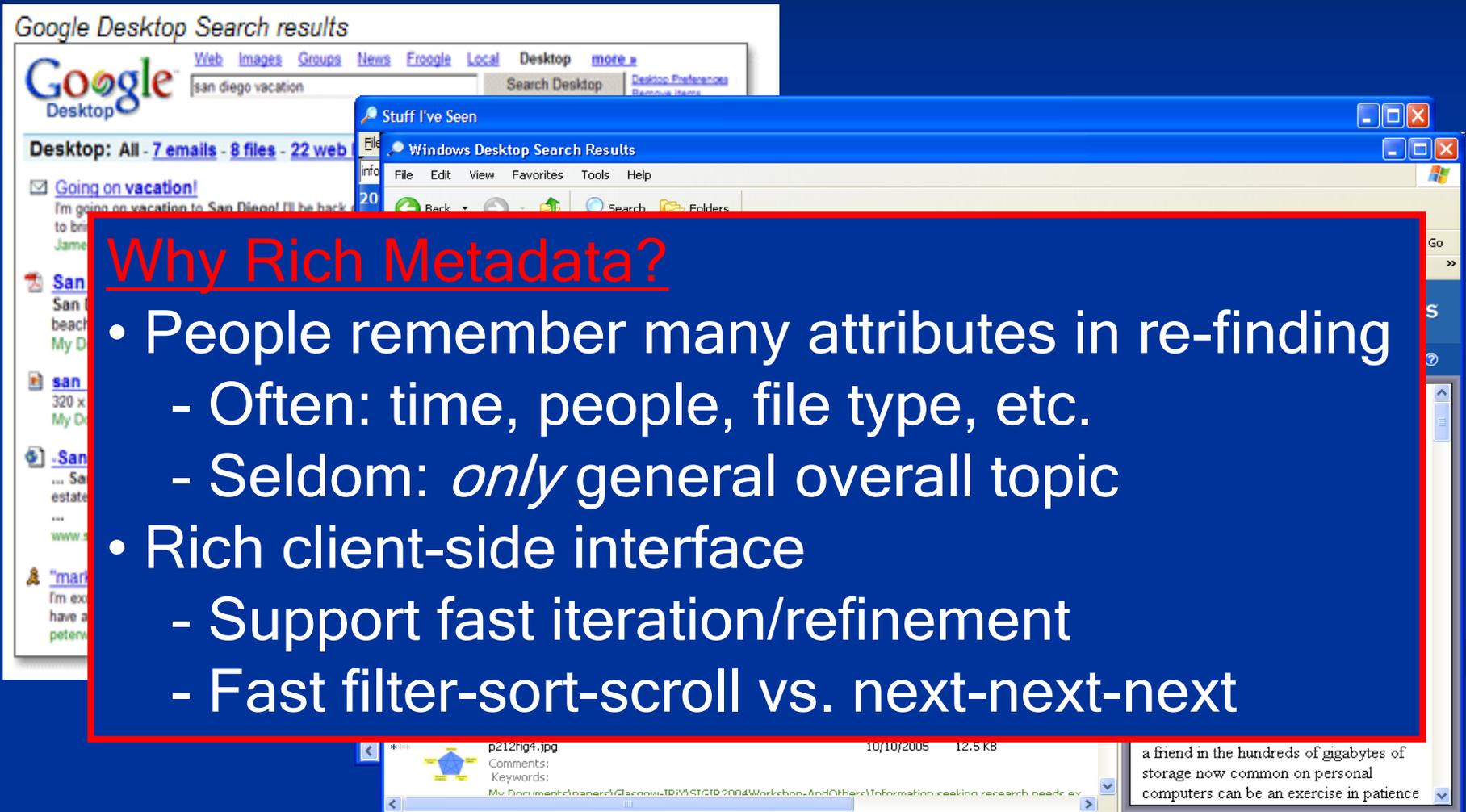


# SIS Usage Data, cont'd

## 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, Author, Artist

# Ranked list vs. Metadata (for personal content)



The image shows a screenshot of a Windows desktop with two search windows open. The background window is 'Google Desktop Search results' showing a search for 'san diego vacation'. The foreground window is 'Windows Desktop Search Results' showing a file named 'p212fig4.jpg' with a date of '10/10/2005' and a size of '12.5 KB'. A blue text box with a red border is overlaid on the search results, containing the text 'Why Rich Metadata?' and a list of bullet points.

## 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

# Re-finding on the Web

- 50-80% page visits are re-visits
- 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%)

# Phlat: Search and Metadata

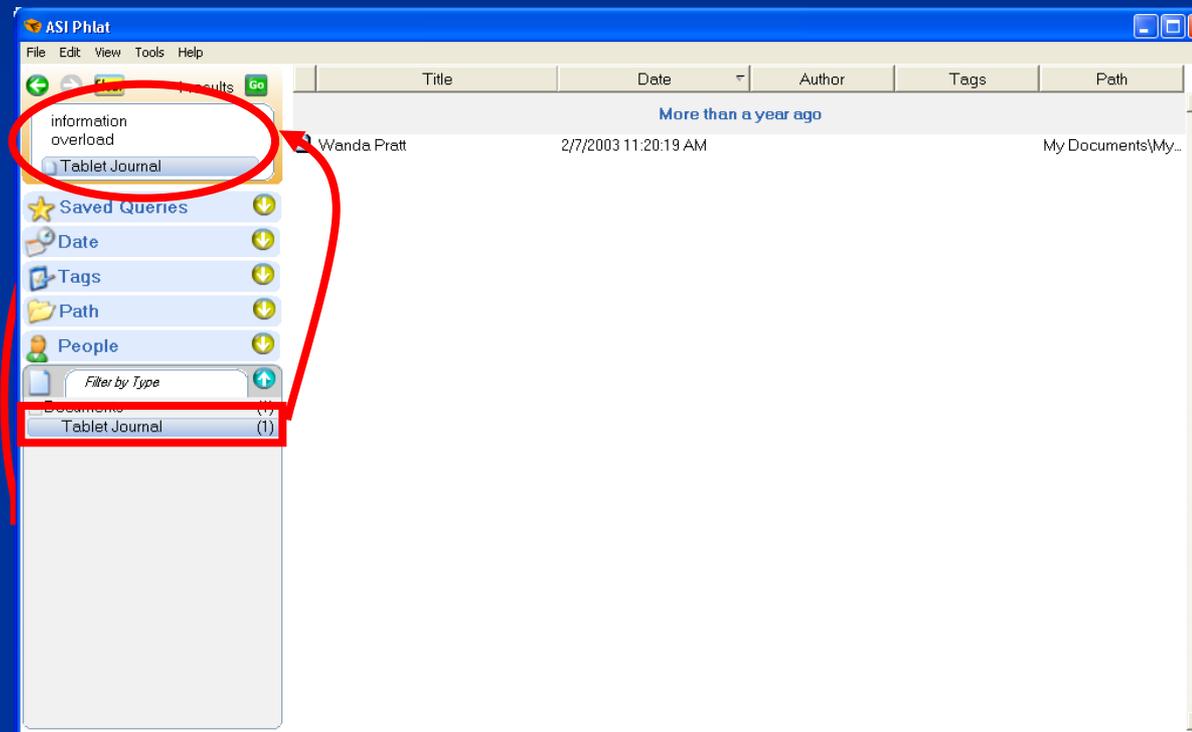
- Phlat (Prototype for Helpful Lookup And Tagging)
  - Shell for WDS; Publically available
  - Tightly couples search and metadata
- Features:
  - Search / Browse (metadata)
  - Unified Tagging
  - In-Context Search

The screenshot displays the Phlat search interface. On the left, the **Query Area** (pink callout) contains a search bar with the text "bike" and a "Go" button. Below it are "Saved Queries" and a "Date" filter. The **Filter Area** (cyan callout) shows a "Filter by Tag" section with a tree view of tags: Personal (123), Family (1), Gaming, Bikes (89), Images (1), Maps (1), Book Clubs (1), Beer, Friends (16), Cats, Art, and New Friends. Below this are filters for Path, People, and Type. On the right, the **Results Area** (red callout) shows a list of search results. The results are organized into sections: "Within the last year" and "More than a year ago". Each result entry includes a checkbox, a title, a date, an author, and a recipient. The first result is "Pedal the Pinchot July 16ht" by Cal Tjader to Willie Bobo, dated 07.05.2005 10:05:0... The second result is "Pedal the Pinchot July 16ht" by Willie Bobo to Cal Tjader, dated 07.05.2005 9:56:00... The third result is "Pedal the Pinchot July 16ht" by Cal Tjader to Willie Bobo, dated 07.05.2005 9:48:20... The fourth result is "Moran State Park Spetember Ride" by Hugh Masekela to Willie Bobo, dated 08.11.2004 9:47:50... The fifth result is "Bye..." by Jimmy Smith to Willie Bobo, dated 08.06.2004 4:42:27... The interface also shows a "Show more results (57)" link.

# Phlat: Faceted metadata

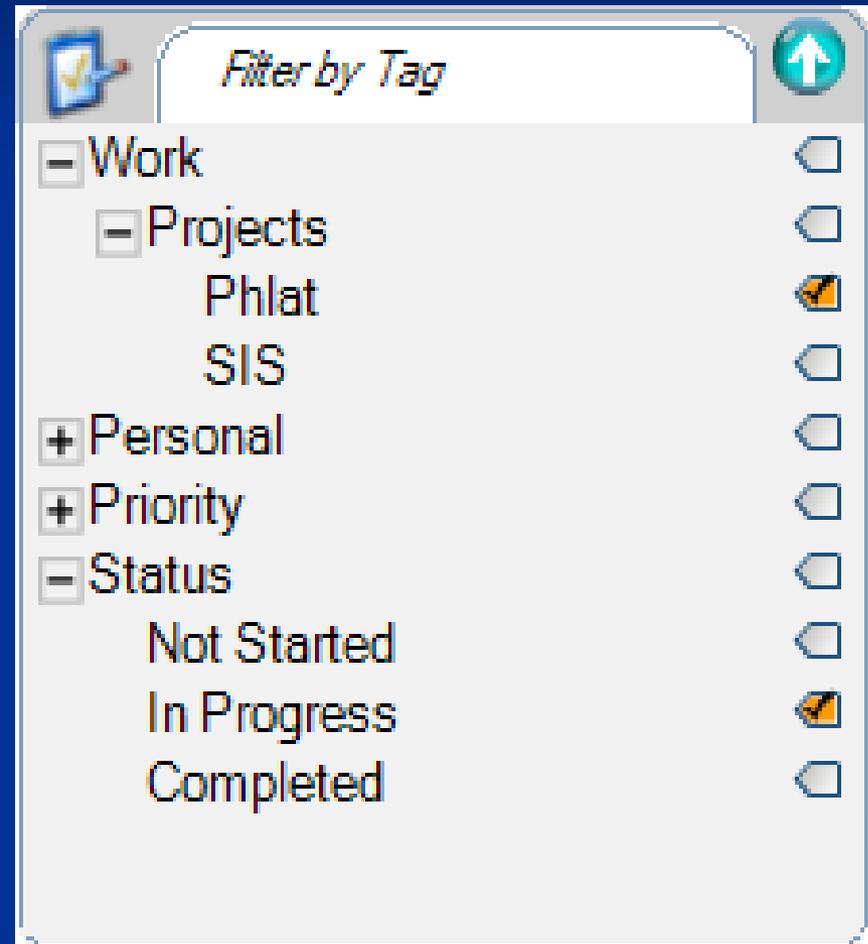
(for filtering, sorting, querying, tagging)

- Tight coupling of search and browsing
- 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.
- “Sideways search”
  - Pivot on metadata
  - Refine or replace query

The screenshot displays the ASI Phlat search interface. The main window shows a list of search results with columns for Title, Date, Author, Tags, and Path. The search query is 'Isa', resulting in 673 results. The interface includes a sidebar with filters for Date, Filter by Tag, Filter by Type, and People. A red circle highlights the 'In Progress' status filter in the 'Filter by Tag' section. The search results list includes items such as 'Susan Dumais Homepage', 'Finkelstein-TOIS2001.pdf', 'kct-proposal', 'ActionLetter-HCI752.R1.pdf', 'Dumais-CV.pdf', 'vita-current.pdf', 'Ise chapter proofs due now.m...', 'lender16.pdf', 'TOP 10 questions from Lenda...', 'Shell triage 8/22', 'vita-current-patents', 'RECAP: Sue Dumais/MaryJo...', 'Briefing: Sue Dumais/MaryJo...', 'How To compare 2 documents', 'MSR/LL @ SIGIR', 'MSR/LL @ SIGIR', 'IOstress Resolutions for 07/...', 'IOstress Resolutions for 07/...', 'Acknowledgment', 'unigram-frequencies.csv', 'ActionLetter-HCI752-R1.pdf', 'ActionLetter-HCI752-R1', 'ActionLetter-R1', and 'Encrypted files/folders (RE...'. The interface also shows a 'Last 30 days' and 'Within the last year' filter, and a 'People' filter section.

# Phlat

The screenshot shows the Phlat search interface. On the left, the 'Query Area' contains a search bar with 'bike' and filters for 'Personal' and 'Mail/Calendar'. Below it is the 'Filter Area' with 'Saved Queries' and 'Date' filters. The 'Results Area' on the right displays a table of search results. The first result is 'Pedal the Pinchot July 16ht' with a date of '07.05.2005 10:05:00...' and author 'Cal Tjader'. The snippet reads: 'Yes, definitely SOMETHING. I think this is twice in a row now that I've had dinner with actually riding with him. —Original Message—'. The path is '/Mailbox' and the tag is '"Personal\Bikes"'. A second result is partially visible below it.

**Query Area**

**Results Area**

**Filter Area**

## 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>**

# Metadata and the Web

See all results that mention **Digital Cameras** New Nikon D200 On Sale

Live Search | MSN | Windows Live | Hotmail United States | Options | cashback | Sign in

Live Search | MSN | Windows Live | Hotmail United States | Options | cashback | Sign in

Live Search seattle

Images 1-12 of 60,900 · [Web](#), [Video](#), [News](#), [Maps](#), [More](#) ▾

SafeSearch Moderate ( [Change](#) )



**Refine by**

- Size
- Aspect Ratio
  - Reset
  - Square
  - Wide
  - Tall**
- Color
  - Reset
  - Black & White
  - Color**
- Style
  - Reset
  - Photograph
  - Illustration**
  - Face

**Related Searches**

- Washington State
- Vancouver
- Portland
- Tacoma

**Scratchpad** New Collection ×

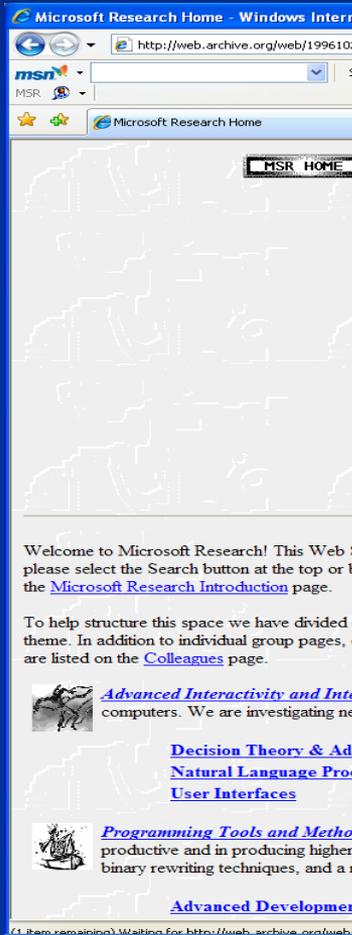
Drag items here to create a new scratchpad collection.

to \$  Go **Nikon D8U - digital camera** by Nikon From **\$924.99**

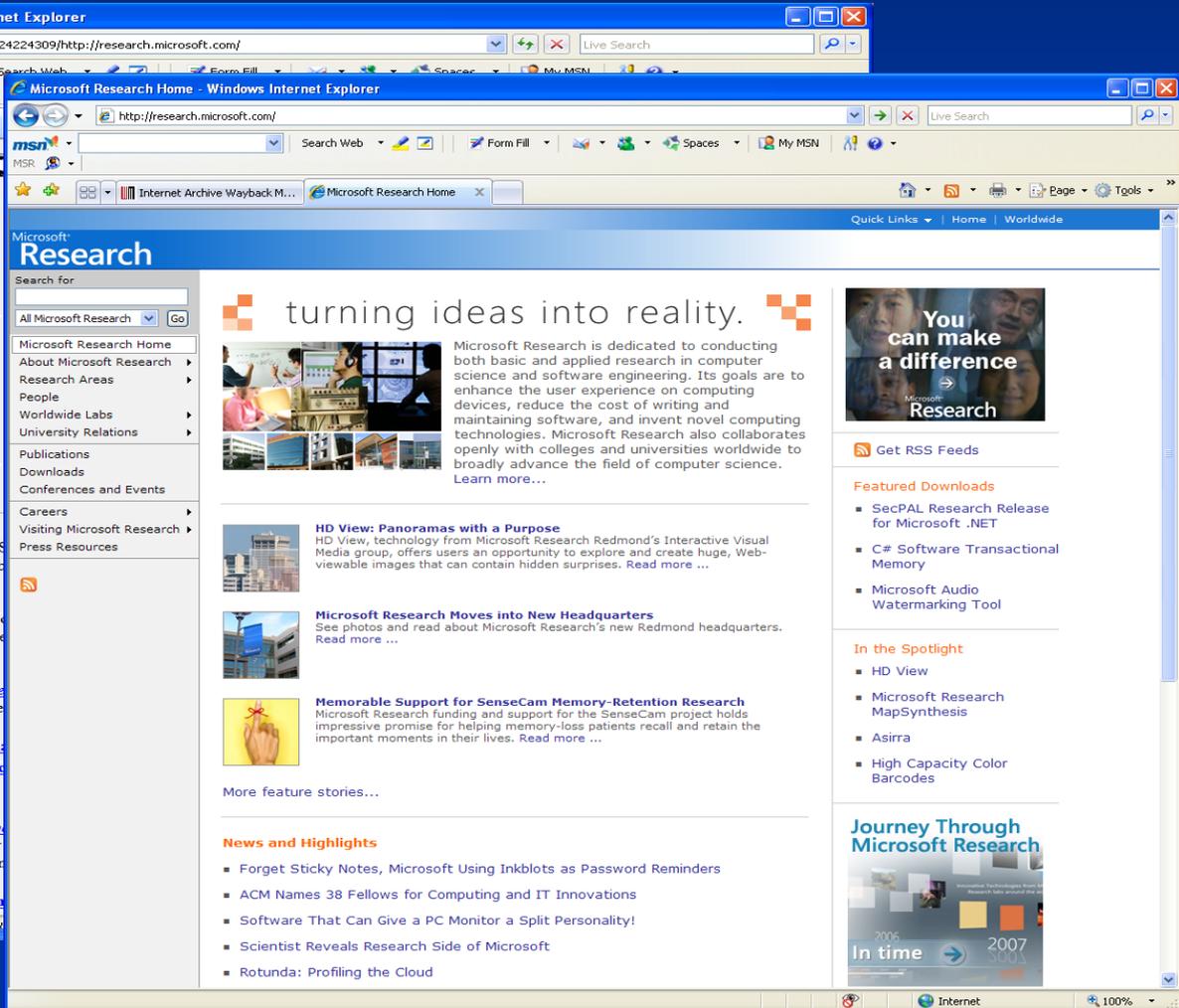
# Dynamic Info Environments

MSR Homepage

1996

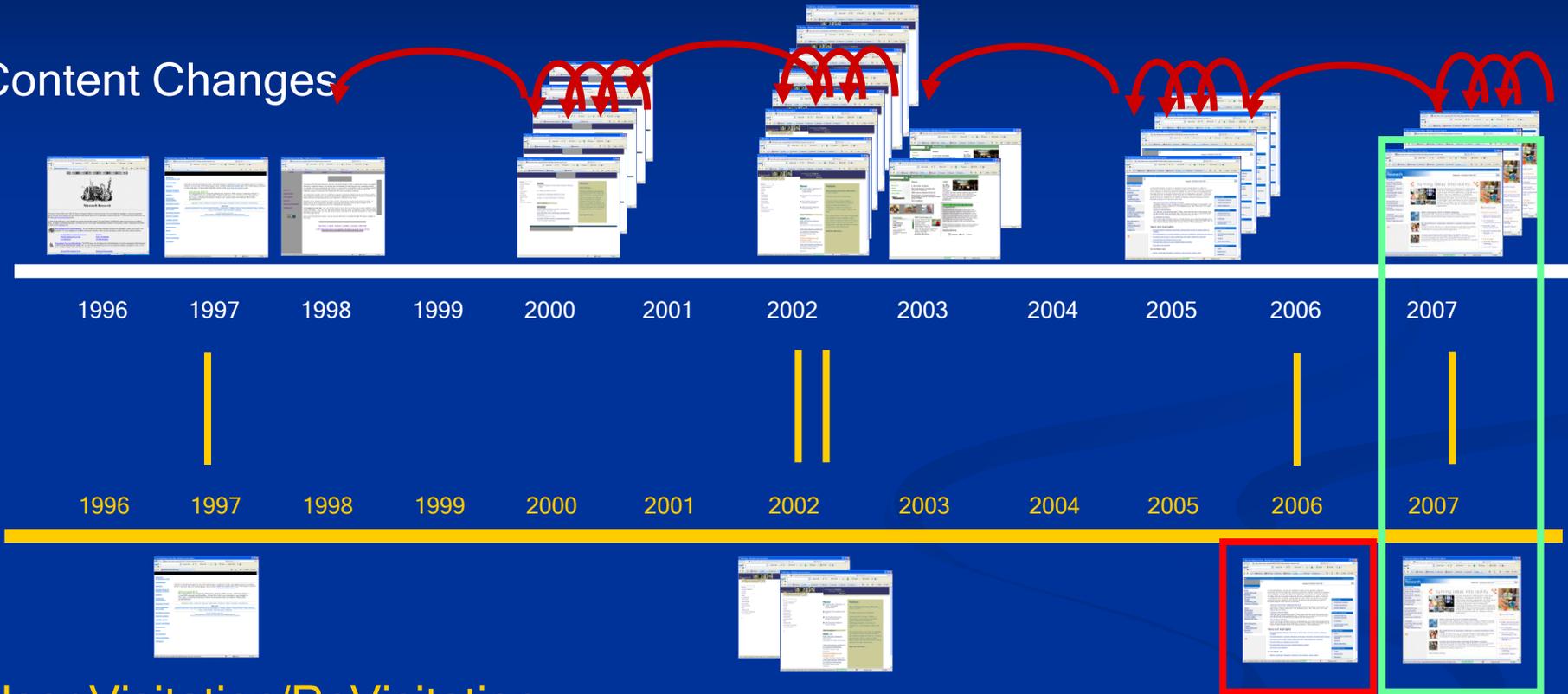


2007



# Dynamic Info Environments

Content Changes



User Visitation/ReVisitation

Today's Browse and Search Experiences

But, ignores ...

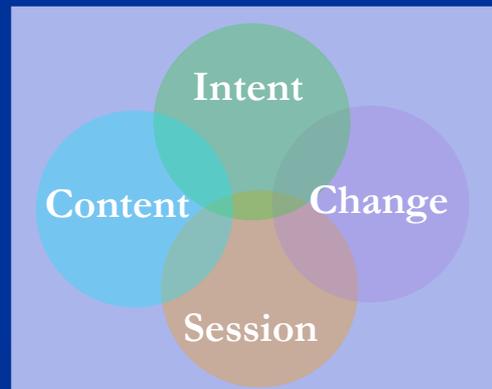
# What We Did

- Content:
  - Crawled 55k pages every hour for 1 year
  - Varying #users, #visits/user, inter-visit interval
- Behavior:
  - Analyzed revisitation patterns for >600k users for these 55k pages
  - Surveyed 20 people for richer understanding of intent
- Examined:
  - **User revisitation** patterns
  - **Page change** patterns
  - Relations between change and revisitation

# What We Found

## Revisitation patterns

- Revisitations to pages are very common
  - 50-80% of pages
- What makes one page's revisits different from another?
- Examined four characteristics



# What We Found

## Change patterns

- 66% of the pages change
  - Change every 123 hours (avg.)
  - Change by 0.21 (avg. dice coeff.)
- Which pages change?
  - Popular pages, .com pages change most
- Which terms change?
  - Term longevity analyses

# What We Found

## Change patterns

1998

Susan Dumais homepage - windows internet explorer

http://web.archive.org/web/19981207032224/research.microsoft.com/~sdumais/

msn

Search Web

Form Fill

Spaces

My MSN

Susan Dumais



Senior Researcher, [Precision Theory & Adaptive Systems Group](#), Microsoft Research

E-mail: [saumais@microsoft.com](mailto:saumais@microsoft.com)

Mail: One Microsoft Way, Redmond WA 98052-6399, USA

**Research Activities:**

I am interested in algorithms and interfaces for improved information retrieval, as well as general issues in and human-computer interaction. I joined Microsoft Research in July 1997. I look forward to working on a wide variety of information access and management issues, including: collaborative filtering, interfaces for improved search and navigation, and user/task modeling. [Stay tuned for new developments here.](#)

Prior to coming to Microsoft, I worked on a statistical method for concept-based retrieval known as Latent Semantic Indexing. You can find pointers to this work on the [Bellcore LSI page](#).

**What's New:**

- Forbes article by William Baldwin on our anti-Spam work. [Spam Killers](#), *Forbes*, Sept 21, 1998, 254-255.
- S. T. Dumais, J. Platt, D. Heckerman and M. Sahami (1998). [Inductive learning algorithms and representations for text](#) appear in: *Proceedings of ACM-CIKM98*, Nov. 1998.
- M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz (1998). [A Bayesian approach to filtering junk e-mail](#). *Proceedings of SIGIR98*, July 27, 1998, Madison, Wisconsin.
- S. T. Dumais (1997). [Tightly coupling structure and search](#). (Powerpoint slides) *SIGIR'97 Workshop on information access*, July 27, 1997, Madison, Wisconsin. A description of the workshop and position papers can be found at [UMass](#).
- S. T. Dumais (1997). [Data mining and the Web](#). (Powerpoint slides) Panel at *Knowledge Discovery and Databases (KDD)*, August 14-17, 1997.

2007

Susan Dumais homepage - windows internet explorer

http://web.archive.org/web/20070206172001/research.microsoft.com/~sdumais/

msn

Search Web

Form Fill

Spaces

My MSN

Susan Dumais



Principal Researcher, [Adaptive Systems & Interaction Group](#), Microsoft Research

E-mail: [saumais@microsoft.com](mailto:saumais@microsoft.com)

Mail: One Microsoft Way, Redmond WA 98052-6399, USA

**We're Hiring at MSR and LiveLabs ...**

We're looking for great folks to advance the state-of-the-art and influence new products in the search arena. We have internships and permanent positions in several areas including: internet search, desktop search, personalization, and novel interfaces for search.

- Intern candidates can apply online at: <http://web.archive.org/web/20070206172001/http://research.microsoft.com/aboutmsr/jobs/internships/default.aspx>
- Permanent employment opportunities at: <http://web.archive.org/web/20070206172001/http://labs.live.com/>

**Research Activities:**

I am interested in algorithms and interfaces for improved information retrieval, as well as general issues in and human-computer interaction. I joined Microsoft Research in July 1997. I work on a wide variety of information access and management issues, including: personal information management, web search, question answering, information retrieval, text categorization, collaborative filtering, interfaces for improved search and navigation, and user/task modeling.

Prior to coming to Microsoft, I worked on a statistical method for concept-based retrieval known as Latent Semantic Indexing. You can find pointers to this work on the [Bellcore \(now Telcordia\) LSI page](#).

**Workshops, Collaborations and Papers:**

- S. T. Dumais (1997). [Data mining and the Web](#). (Powerpoint slides) Panel at *Knowledge Discovery and Databases (KDD)*, August 14-17, 1997.

# What We Found

## Change patterns - rate of change

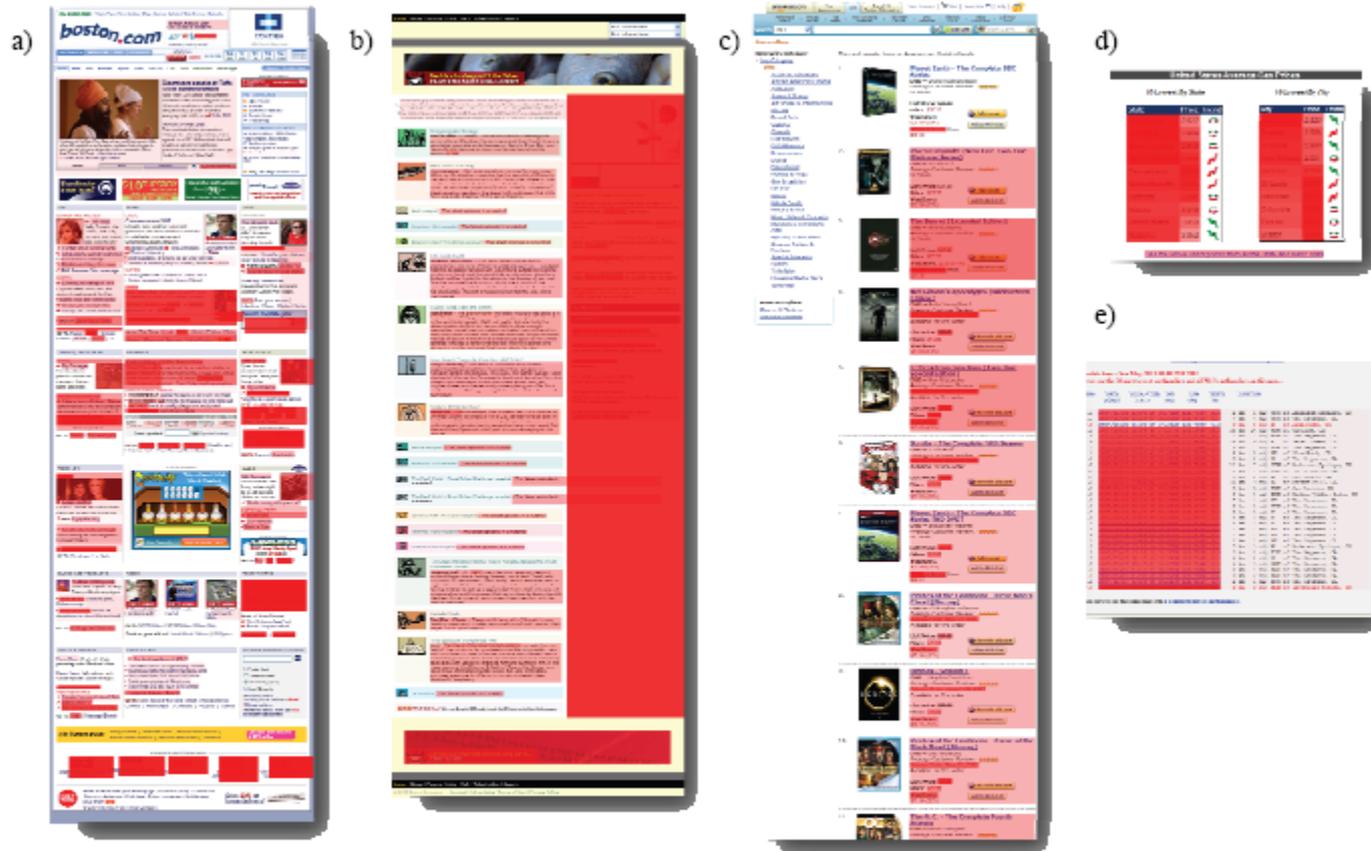


Figure 7. Renderings of the lifespan of elements on a number of pages (darker red blocks are shorter life spans) including a) boston.com, b) televisionwithoutpity.com (note the groups of similarly colored content), c) the DVD bestseller list on Amazon, d) gas prices in various cities on GasBuddy.com, and e) a list of earthquakes at the USGS. Not all blocks marked.

# What We Found

## Change patterns - for your visits

Diff-IE

File Edit View Favorites Tools Help

msn MSR

Search Web Form Fill

Diff-IE abc Compare to 15 min ago Load Options

Seattle news, sports, entertainment | seattlepi.com

MomSeattle · DadSeattle · SPI · MySeattlePix · My account

seattlepi.com

64° 5-day forecast

Our network sites seattlepi.com help

Local US/World Sports Business A&E Life Comics Photos Opinion Blogs Subscribe Buy Ads Jobs Autos Homes Rentals NWsource

Tuesday, April 29, 2008

**Seattle home prices drop again**  
Seattle-area annual appreciation has now declined for 26 straight months, and the value is down 6.5 percent from its peak in July 2007. **New**

**Obama says he's outraged by former pastor's comments** **New**

**Was \$1 billion a timber slush fund or fair deal?**

**Tax rebate 'splurge' on gas, food and bills**

**Alleged Jimi Hendrix sex tape to be released** **New**

**FEATURED GALLERIES**

Sports Sideshow Star Sightings Australian Fashion Week

More Galleries | MySeattlePix | You've Been Spied

**SCOREBOARD**

4:05 PM Seattle Cleveland

Mariners scoreboard

ADVERTISING

**DSL & Home Phone plans start at**

**\$49.95** a month

**with your first month FREE!**

**ADVERTISING**

**Fry's** click to see this week's specials >

**mypi** Personalize Now

ap: top headlines

- Obama says he's outraged by former pastor's comments
- Bush faults Congress for inaction on energy, housing bills
- Officials: Captive family in Austria reunited
- Concentration camp doctor heads list of top 10 wanted Nazis
- Army widens probe after finding bad conditions at Fort Bragg

» more

ap: odd news

- Wisconsin man starts Web site devoted to roadkill
- Family arrested on drug charges after reporting fight
- 10-year-old monkey at Utah zoo gets artificial lenses
- Saudi governor orders haircuts for men who hit on women
- Man accused of T-shirt thefts as girlfriend applies for job

» more

most read

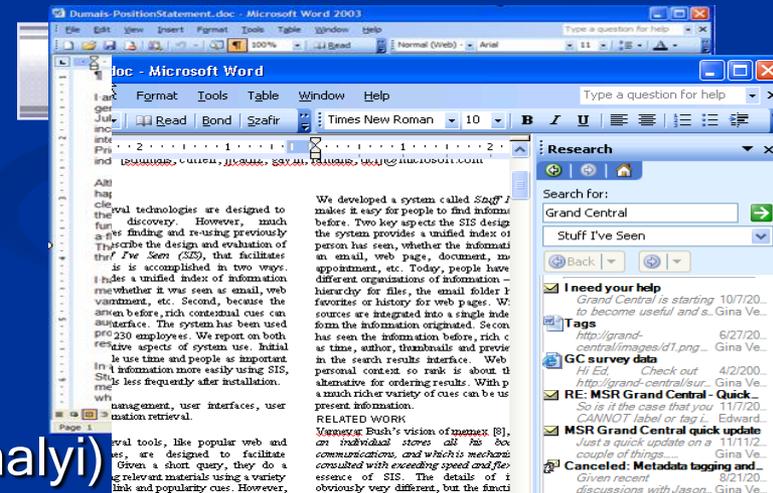
- Stabbed over TV 'Top Model' noise
- Obama says he's outraged by former pastor's comments
- Seattle house prices lowest since summer '06
- Alleged Jimi Hendrix sex tape to be released
- Mariners appear ready to call up Jeff Clement

» more

# Search in Task Contexts

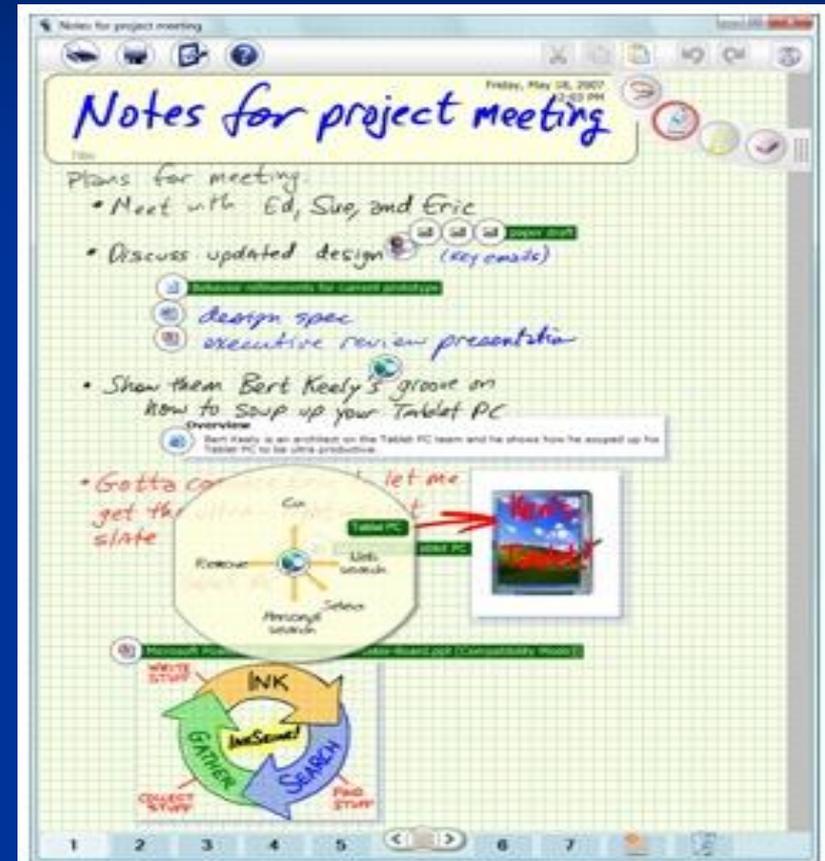
- Search is not the end goal ...
- Support information access in the context of ongoing activities (e.g., writing talk, finding out about, planning trip, buying, monitoring, etc.)

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



# *InkSeine*: Active Note Taking

- Tablet application for active note taking
- Unifies ink, search and gather functions into a fluid workflow
- Note taking, enriched w/:
  - Search from ink
  - Show results in app
  - Integrate results, links and clippings into notes
  - Maintain work flow
- “Inking for thinking”



An example of an InkSeine note with links, searches, and screen clippings.

# 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
  - Context search, Notes, Chat, Tags, Inlinks, Queries
  - <http://www.communitybar.net>
- Implicit Queries (IQ)
  - Also Y!Q, Rememberance Agent, Watson, Query-free search
- Even more possibilities for context-driven retrieval w/ rich sensors and ubiquitous networks

The screenshot shows a Microsoft Internet Explorer browser window. The address bar displays 'http://msn/CommunityBar/'. The main content area features a search bar with 'Community' and 'Web' options, a 'SEARCH' button, and a 'GO' button. Below the search bar, there are several notes and a chat window. The chat window shows a message from 'Richardson' and a response from 'Brill\_0322'. The website background is 'jazzalley.com' with a 'Dimitriou's jazz alley Restaurant & Nightclub' logo and a navigation menu including 'Home', 'Calendar', 'Club History', 'Dining', 'General Info', 'Directions', 'Reservations', 'PJI Info', 'Mailing List', 'Jobs', and 'Hot Links'. There are also advertisements for 'Online Shop' and 'Photo Gallery'.

# Documents as Context (Implicit Query, IQ)

- Proactively find info relevant to item being read/created
  - Quick links
  - Matching content (several sources)
- Challenges
  - Relevance, ok
  - When to show? (useful)
  - How to show? (peripheral awareness)

The screenshot shows an email client window titled "I need your help - Message (HTML)". The email content is as follows:

From: Gina Venolia  
To: ASI and Affiliates  
Cc:  
Subject: I need your help

Sent: Tue 10/7/2003 10:08 AM

Grand Central is starting to become useful and stable enough to "dogfood". Yay! To make sure that it works on a variety of machine configurations, I'm hoping that I can get some volunteers to run Grand Central on their machines. Requirements and installation steps are at <http://grand-central/install.html>

Thanks,  
-Gina  
Gina Venolia

Callouts from the image:

- Quick links for People and Subject.** (points to the "Search SIS for:" section in the IQPane)
- Background search on top  $k$  terms, based on user's index —**  
Score =  $tf_{doc} / \log(tf_{corpus} + 1)$  (points to the email content)
- Top matches for this Implicit Query (IQ).** (points to the "Subj: I need your help" in the IQPane)

# *P*Search: Personalized Search (Even Richer Context)

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

Step 1: retrieve >> 10 results

Live Search ac1 2008?

Web 1-10 of 3,480,000 results - [Advanced](#)  
See also: [Images](#), [Video](#), [News](#), [Maps](#), [More](#)

**Personalized** [We](#) [Cache](#) [Desktop](#)

- [ACL-08: HLT - Home Page](#)  
o <http://www.ling.ohio-state.edu/acl08/>
- [ACL 2008 Workshop on Mobile Language Processing](#)  
o <http://www.mobilenlpworkshop.org/>
- [ACL 2008 Lineup Announced! Sonic Itch Music](#)  
o <http://www.sonicitchmusic.com/?p=1863>

...show more

[Austin City Limits Music Festival : September 26-28, 2008 : Austin, TX ...](#)  
and <sup>TM</sup> designate trademarks of Capital of Texas Public Telecommunications Council, Inc. All Rights Reserved.  
[www.aclfestival.com/default.aspx](http://www.aclfestival.com/default.aspx) · [Cached page](#)

[Austin City Limits - Lineup](#)  
and <sup>TM</sup> designate trademarks of Capital of Texas Public Telecommunications Council, Inc. All Rights Reserved.  
[www.aclfestival.com/lineup.aspx](http://www.aclfestival.com/lineup.aspx) · [Cached page](#)  
[Show more results from www.aclfestival.com](#)

**ACL Connections 2008**  
Connections 2008 was a resounding success. Attendees learned new and innovative ways to use ACL technology and got a sneak peak at AuditExchange 2008, including live demonstrations  
[www.acl.com/userconference](http://www.acl.com/userconference) · [Cached page](#)

(user model)

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

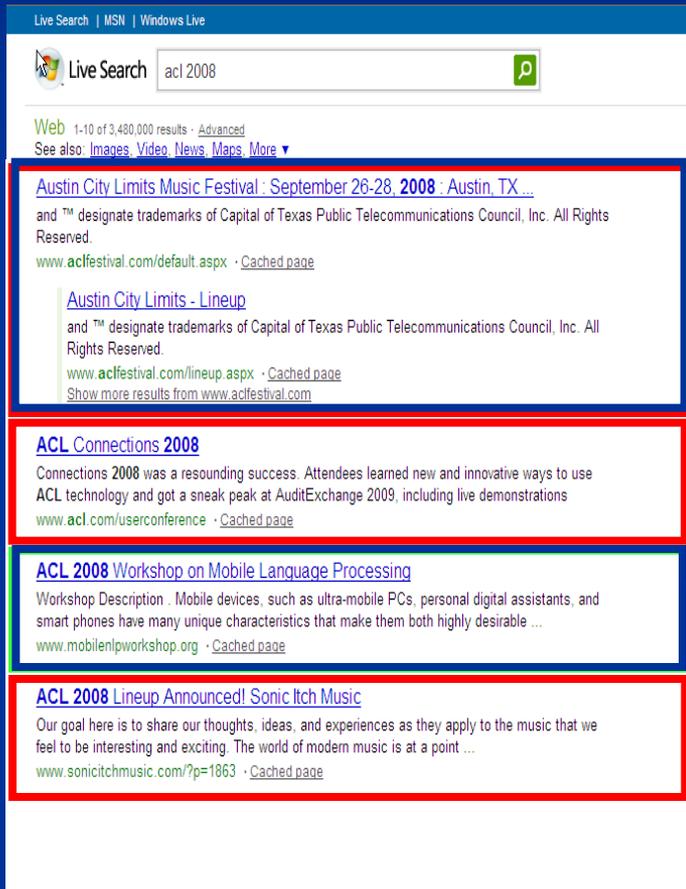


# Building a User Profile

*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 =  $f(\text{Cont}, \text{Beh}, \text{Web})$

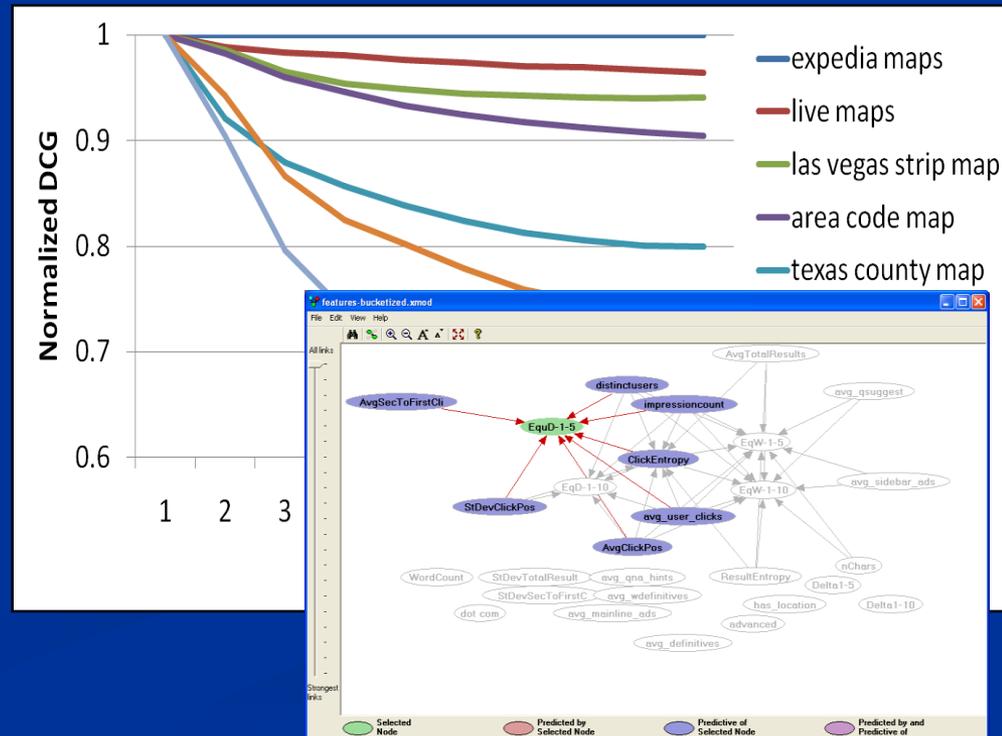
■ P\_Content Match:  
 $\text{sim}(\text{result}, \text{user\_content\_profile})$

■ P\_Behavior Match:  
 $\text{visited URLs and sites}$

■ 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)*”



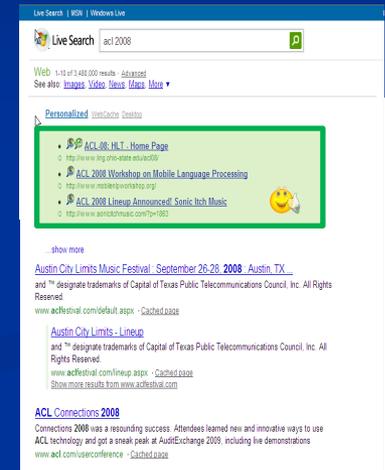
# How to Personalize Display

## ■ Presenting results

- Inline display (for demo)
  - Also: tabs, slider, fisheye, metadata
- Interleave results (for evaluation)
- Behind the scenes (for the curious)
- Balance consistency, novelty

## ■ Summarizing results

- Highlight results that were seen before
- Highlight new result content
- Personalized snippets



[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](#)

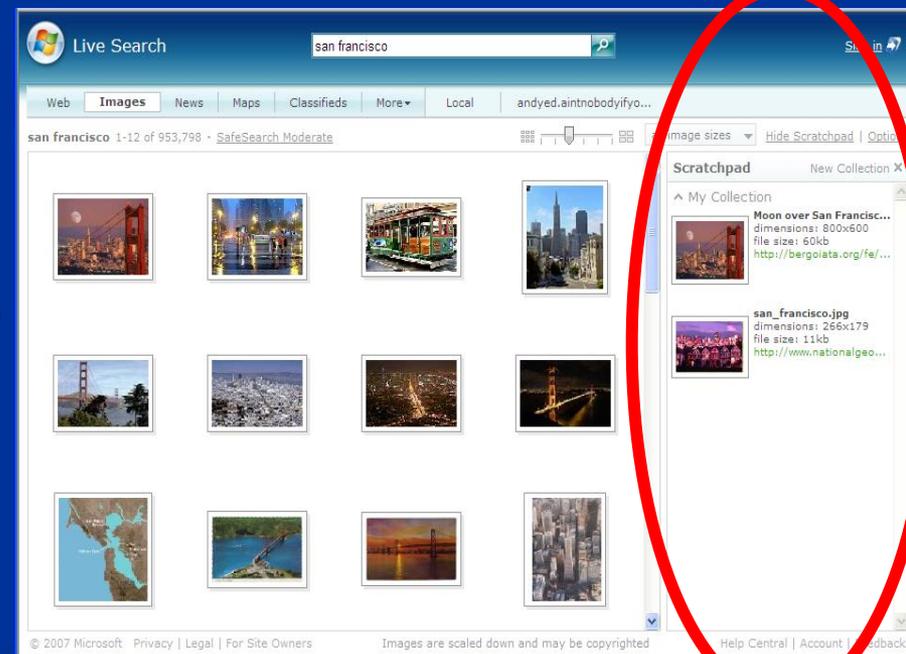
# More “Personalized” Search

- PSearch - rich long-term context; single individual
- Short-term session/task content
  - Query: *ACL*, ambiguous in isolation
    - austin music ... tickets alison krauss ... *ACL*
    - natural language processing ... summarization ... *ACL*
    - knee surgery ... orthopedic surgeon ... *ACL*
- Groups of similar people
  - Groups: Location, demographics, interests, behavior, etc.
    - Freyne & Smyth (2006); Smyth (2007); Teevan & Morris (2008)
  - Mei & Church (2008)
    - $H(\text{URL}) = 22.4$
    - Search:  $H(\text{URL}|\text{Q}) = 2.8$
    - “Personalization”:  $H(\text{URL}|\text{Q}, \text{IP}) = 1.2$
  - Many models ... smooth individual, group, global models

# Beyond Search - Gathering Info

- Support for more than “retrieving” documents
  - Analyze -> Use -> Share
  - Exploratory search
- Lightweight scratchpad or workspace support
  - Iterative and evolving nature of search
  - Resuming at a later time or on other device
  - Sharing with others

ScratchPad



# 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)

SearchTogether



Figure 1. The SearchTogether client. (a) integrating messaging, (b) query awareness, (c) current results, (d) recommendation queue, (e)(f)(g) search buttons, (h) toolbar, (i) browser

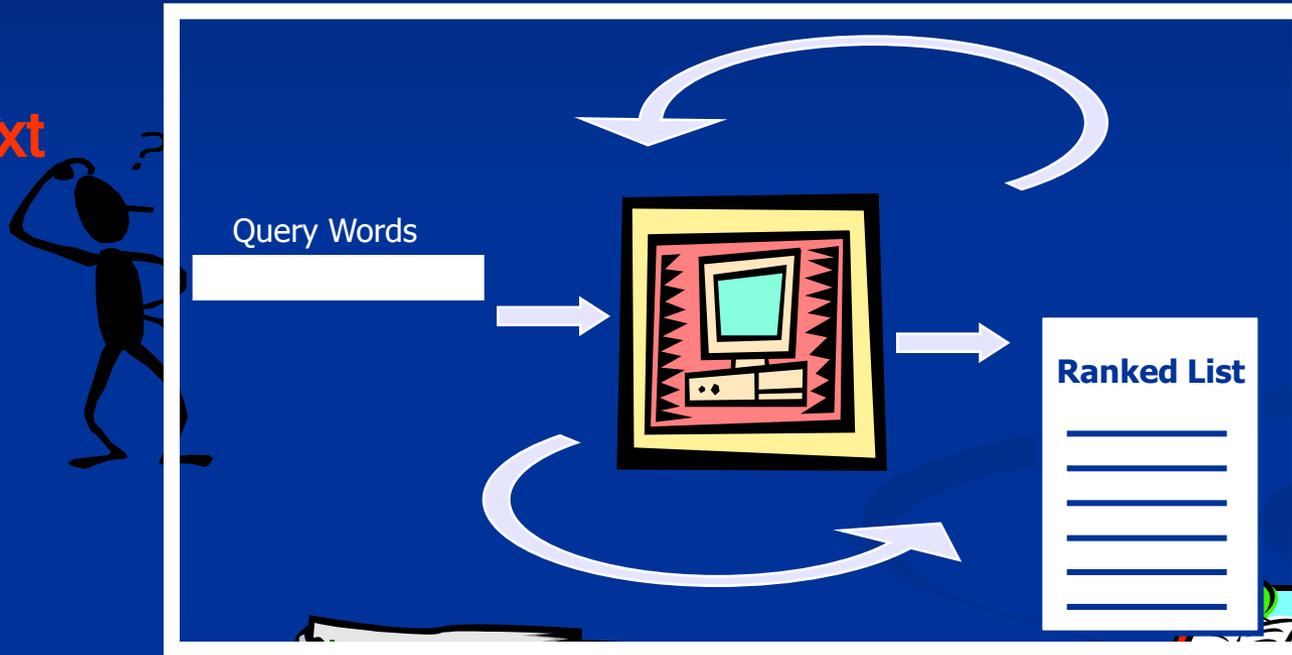
Download: <http://research.microsoft.com/searchtogether>

# Looking Ahead ...

- Continued advances in scale of systems, diversity of resources and quality of ranking, etc.
- Tremendous new opportunities to support information retrieval and analysis by ...
  - Understanding user intent
    - Representing non-content attributes and relations
    - Modeling user interests and activities over time
  - Supporting the search process
    - Developing interaction and presentation techniques that allow people to better express their information needs
    - Supporting analysis, use and sharing of results
  - Considering search as part of richer landscape

# Thinking Outside the (Search) Box

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>
- InkSeine, <http://research.microsoft.com/InkSeine>
- Search Together, <http://research.microsoft.com/searchtogether>