

First and Seventeenth International Conference on User Modeling, Adaptation, and Personalization

## Thinking Outside the (Search) Box

Trento, Italy 22-26 June 2009

Susan Dumais Microsoft Research http://research.microsoft.com/~sdumais

## Outline

Search Yesterday and Today Search and Context User, Domain and Task Contexts Examples of Prototypes w/ Richer Context Stuff I've Seen (desktop) Personalized Search (web) ■ Etc.

Challenges and Opportunities

## 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<sup>1</sup>0 ...

#### 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

Search Search the web using Google! Google Search I'm feeling lucky Search	
Search the Web:	
Search	
Web   MSN   Images   Videos   Shopping   News   Maps   NEW	1
Ding	b
	Go
Search the Web Images Groups Directory News	
Web search results         L+16 of 30% results mod relevant to edante ehabitet         Model 5 and the top 1 into summaria 1 Sort Votate 1 Unitopy results         Model 5 and the top 1 into summaria 1 Sort Votate 1 Unitopy results         Bury the book         Search 5 and the top 1 into summaria 1 Sort Votate 1 Unitopy results         Search 5 and 5 an	ch
A weets, rhu, sue r, Perry Ha viteto 1995 - 1997 way 1997 reactional biological bence, Leetown Vin so, Idad Science Center, W.V. Dinision of Natural Resources 1. Quartify microhabitat use for Star Wars must-have at 20% Date 9 Jan 1998 Sara 39% http://www.carkwited/coops/altade/html	
BAYOU DARTER, Etheostena (Nothenotes) rahmin U.S. Eth. I. Wildlife Sergice Source: PNI Fagen 4 - Ace 3 201 Percenta I handanic throughout dramgs, Experimer 22, 1375 1915, Date: 13 Art 1996, Sec 11. (His/Jown ace/Machings/Machina Tahmi	Encarta Search
Eind similar pages 1 Geoped results form www.lws.giv CWT.Witherbeard D. ALL RESULTS 1-10 of 1,390,000,000 results - <u>advanced</u> Cheater Welfer Tra Web Pages 1997. BYD. Direct Sale 12 - A hourt Mich December Logan Upun To December The Mich of About Mich December Call	New
Bits         Search         Earn How To Search The Web at About Web Search           Search         Web search secrets, search engine information, Web 2.0 tools and tips and more at About Web           Search         your best resource for learning how to search the Web more productively.	Search
websearch.about.com - <u>cached page</u> We Web search engine - Wikipedia, the free encyclopedia	
A Web search engine is a tool designed to search for information on the World Wide Web. The search results are usually presented in a list and are commonly called hits. The information may consist of web pages, images, information and other types of files. Some search engines also mine	Web
data available in databases or open directories. We en.wikipedia.org/wiki/Web_search_engine : <u>cached page</u>	
Google Search I'm Feeling Lucky	Uive Search
Advanced Search Parteriose	
Google Search I'm Feeling Lucky	<u> </u>

## **Supporting Searchers**

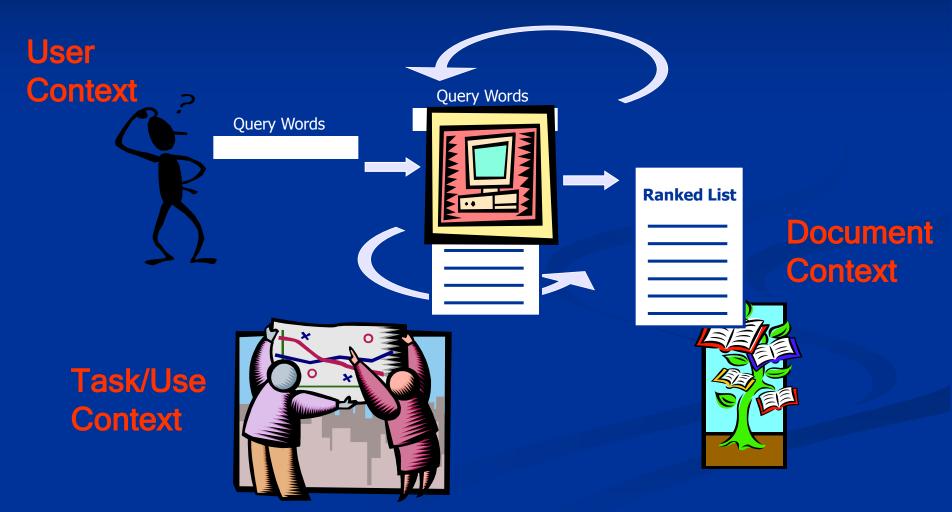
- The search box
- Spelling suggestions
- Query suggestions
- Auto complete
- Inline answers
- Richer snippets

But, we can do be ... by using contents



Q

## Seasea and Todayext



## **Search and Context**

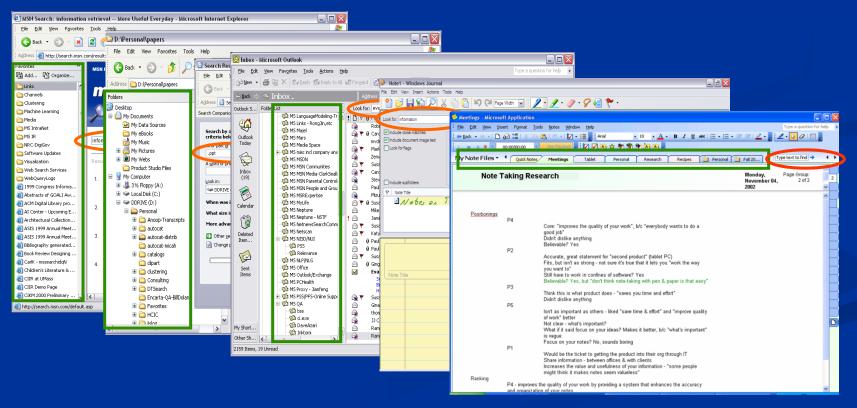
Research prototypes: provide insights about algorithmic, user experience, and policy challenges

#### 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 (uRank, SearchTogether)

## **Information Solos**

Many sources of information, and methods for finding and organization (e.g., web, mail, contacts, docs, photos, notes)



Dumais et al., SIGIR 2003

## 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
- Re-finding vs. finding

 Vista Desktop Search (and XP, Live Toolbar)
 Also, Spotlight, GDS, X1, ...
 UMAP, June 24 2009

Ele View Tools Help			O6	C 17	0
nformation overload		- D	Go Exact Match	rive	Seen
200 rows returned					
Document	\land Date	Rank	Paul	Author	Mail 10
✓ (Al) (200)	🗹 (AI) (200)	-		-	-
Web Pages (77)	Today (0)	_			_
<ul> <li>Outlook (72)</li> </ul>	Yesterday (0)				
+ ¥ Fies (51)	Last 7 days (1)				
	Last 30 days (.				
	Older than 30				
	5/6/2004 11:27 2		Last 7 days personal folders/media	Km Davis	Susan Dumais
Briefing: AP/Sue Dumais 5/6@4:30 p.m.	5/6/2004 11:272	8	Last 30 dava	Km Davis	Susan Dumais
ArFinal Recap: The Economist Campus Visit April 15	4/20/2004 6:21 2	7	Last 30 days personal folders/media	Kinten Birkeland	Karen Bedetzki: Suzan DelBene: Jo
			Older than 30 days		
Microsoft: Longhom to arrive in 2005 - News - ZD	3/26/2004 10:2 6		temporary internet files\content.ie5\wty7opub		
News: All about Longhom	3/26/2004 10:2 8		temporary internet files\content.ie5\0xubghir		
X1 instantly searches files & email. For Outlook, Ou			temporary internet files\content.ie5\z79fh0w		
Enhanced Microsoft Exhibit Offers Sneak Peek at			temporary internet files\content.ie5\y5nshw7a		
d prefaceforprogram	3/22/2004 5:011	4	personal folders/hit-naacl04		
prefaceforprogram	3/22/2004 1:491		d:\personal\papers\reviews\ht-naacl04	sdumais	
HLT/NAACL preface	3/22/2004 8:181		personal folders/hit-naacl04	Salim Roukos	marcu@ISI.EDU; Susan Dumais
The Future of Information Filtering	3/20/2004 12:0 2		temporary internet files\content.ie5\szfn6ob		
Slashdot   Google buys Pyra Labs	3/17/2004 3:221		temporary internet files\content.ie5\hko3hds5		
Recap: Steven Levy/Newsweek on IQ	3/15/2004 6:277		personal folders/media	Peter Wootton	Eric Horvitz
Homepage for HLT-NAACL 2003	3/15/2004 5:342		temporary internet files\content.ie5\ep381kvu		
2 pc301_scotthay[1]	3/12/2004 10:4 4		temporary internet files\content.ie5\0zbfu05t	Scott Hay	
Pedro Domingos	3/10/2004 4:001		temporary internet files\content.ie5\xkkr158h		
JEP: Multimedia Features and Information Retrieval			temporary internet files/content.ie5/z79ffh0w	author name	
Bates' Bibliography: Information Seeking, Indexing	3/8/2004 3:54 2	0	temporary internet files/content.ie5/ep381kvu		

🕞 🕤 🕫 ) Search Results in Indexed L	ocations				Windows DS
Show only: A E-mail Document Pi	cture Music Oth	er			vvinuows DS
Nganize 👻 🏢 Views 💌 📠 Preview	r 👻 🔜 Save Search	📄 Search Tools 👻 🚔 P	rint 👿 Share 🕚 Burn		
Name	Date modified	Туре	Folder	Authors	Tags ^
RE: FW:	10/3/2008 8:07 AM	Outlook E-Mail Message	Sent Items (Mailbox - Susan D	Susan Dumais	
<ul> <li>DeliaReceipt.txt</li> </ul>	10/2/2008 9:01 PM	Shortcut	Recent Items (C:\Users\sdum		
Aug2007-DeliaBenVi.jpg	9/14/2008 3:13 PM	JPEG Image	Moms90th - Copy (C:\Users\s		
🔤 Aug2007-DeliaBenVi.jpg	9/14/2008 3:13 PM	JPEG Image	2008-05May-Calif (C:\Users\s		
Aug2007-DeliaBenVi.jpg	9/14/2008 3:13 PM	JPEG Image	PartyPictures (C:\Users\sdum		
RE: gift wrap fundraiser for Summit Spee	9/11/2008 8:58 AM	Outlook E-Mail Message	Sent Items (Mailbox - Susan D	Susan Dumais	
View my photos: Delia's school at the zoo	5/26/2008 5:26 PM	Outlook E-Mail Message	MiscMail/MichaelPreston (Pe	mdpreston@gmail.	
RE: A blast from the past	5/7/2008 10:39 PM	Outlook E-Mail Message	MiscMail (Personal Folders 20	Susan Dumais	
View my photos: April special events	4/29/2008 8:19 PM	Outlook E-Mail Message	MiscMail/MichaelPreston (Pe	mdpreston@gmail.	
🗎 Re: Sorry	1/21/2008 7:15 PM	Outlook E-Mail Message	MiscMail/MichaelPreston (Pe	Michael Preston	
G Sorry	1/21/2008 6:35 PM	Outlook E-Mail Message	MiscMail/MichaelPreston (Pe	Michael Preston	Carrier Carrier
🖻 scooter for Delia	11/15/2007 9:58 AM	Outlook E-Mail Message	ImptMiscPersonal/Xmas2007	Christine Preston	and the second s
11 Day Report 06-12-07.xlsx (Action: Prep	6/1/2007 9:27 AM	Microsoft Office Excel Wo	Candidates/Interns (Personal	v-jesiel	
raw and some analysis-STDMods.als	8/25/2006 10:07 AM	Microsoft Office Excel 97	20060609 (C:\Users\sdumais\		
pS8-cutrell.pdf	8/10/2006 9:59 AM	Adobe Acrobat Document	MiscPapers (C:\Users\sdumai		
58-cutrell.pdf	8/10/2006 9:59 AM	Adobe Acrobat Document	Misc (C:\Users\sdumais\Docu		
Sexport4.csv	6/14/2006 7:20 AM	Microsoft Office Excel Co	20060606 (C:\Users\sdumais\		
Sclick-vs-webclick.csv	6/9/2006 3:51 PM	Microsoft Office Excel Co	20060609 (C:\Users\sdumais\		
Click-vs-noclick.csv	6/9/2006 3:22 PM	Microsoft Office Excel Co	20060609 (C:\Users\sdumais\		
sport - excluding control group.csv	6/9/2006 2:44 PM	Microsoft Office Excel Co	20060609 (C:\Users\sdumais\		
seport - removed control and web befor	6/9/2006 2:18 PM	Microsoft Office Excel Co	exclude_before_click (C:\User		
Sigram-frequencies.csv	6/9/2006 1:16 PM	Microsoft Office Excel Co	20060609 (C:\Users\sdumais\		
Sexport.csv	6/6/2006 5:50 PM	Microsoft Office Excel Co	20060605 (C:\Users\sdumais\		
🔁 export3.csv	6/6/2006 3:44 PM	Microsoft Office Excel Co	20060606 (C:\Users\sdumais\		
Sexport2.csv	6/6/2006 11:00 AM	Microsoft Office Excel Co	20060606 (C:\Users\sdumais\		
1 20060203_all.csv	6/2/2006 4:08 PM	Microsoft Office Excel Co	20060203 (C:\Users\sdumais\		
< l					•
Aug2007-DeliaBenVi.jpg JPEG Image	Date taken: <b>Specify da</b> Tags: Add a tag Rating: 습습감	te and time t Dimensions: 4 Size: 3 なな Title: A	1.0 KB Comr	thors: Add an author nents: Add commen naker: Add text	

## SIS Demo

₽ Stuff I've Seen							
<u> Eile V</u> iew Tools <u>H</u> elp							
information overload		- D	Go Exact Match 💌 Clear All )				
200 rows returned							
Document	Date	Rank	Path		Author	Mail To	
✓ (All) (200)	✓ (All) (200)	-		-	-	-	-
✓ Web Pages (77)	Today <b>(0)</b>	_		_			_
+ ✓ Outlook (72)	Yesterday (0)						
+ ✓ Files (51)	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 11:272	8	personal folders/media	۲	ûm Davis	Susan Dumais	
Image: Provide the April 15	A /20 /2004 C-21 2	7	Last 30 days personal folders/media	L	ûrsten Birkeland	Karen Redetzki: Suzan DelBene:	lebe
Campus Visit April 15		/	Older than 30 days		VISIEN DIREIGNU	Nateri nedelzki, Suzari Delberie, i	JOHN
Microsoft: Longhom to arrive in 2005 - News - ZD	3/26/2004 10:2 6		temporary internet files\content.ie5\wty7opub				_
News: All about Longhom	3/26/2004 10:28		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			temporary internet files\content.ie5\y5nshw7a				
prefaceforprogram	3/22/2004 5:011		personal folders/hlt-naacl04				
prefaceforprogram	3/22/2004 1:491	4	d:\personal\papers\reviews\hlt-naacl04	s	dumais		
HLT/NAACL preface	3/22/2004 8:181	4	personal folders/hlt-naacl04	5	alim Roukos	marcu@ISI.EDU; Susan Dumais	
The Future of Information Filtering	3/20/2004 12:0 2	5	temporary internet files\content.ie5\szfn6ob				
🔄 Slashdot   Google buys Pyra Labs	3/17/2004 3:221	5	temporary internet files\content.ie5\hko3hds5				
Recap: Steven Levy/Newsweek on IQ	3/15/2004 6:277		personal folders/media	F	eter Wootton	Eric Horvitz	
Homepage for HLT-NAACL 2003	3/15/2004 5:342	-	temporary internet files\content.ie5\ep381kvu				
Dc301_scotthay[1]	3/12/2004 10:4 4	-	temporary internet files\content.ie5\0zbfu05t	5	Cott Hay		
Pedro Domingos	3/10/2004 4:001		temporary internet files\content.ie5\xkkr158h				
JEP: Multimedia Features and Information Retrieval			temporary internet files\content.ie5\z79ffh0w	a	uthor name		
Bates' Bibliography: Information Seeking, Indexing,	3/8/2004 3:54 2	0	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

#### Query characteristics

- Short queries (1.6 words)
- Few advanced operators or fielded search in query box (~7%)
- But ... 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					
Total	91k items	25.4 Gb					
Index		190 Mb					
		+1.5 Mb/week					

## 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

Log(Freq) = -0.68 \* log(DaysSinceSeen) + 2.02



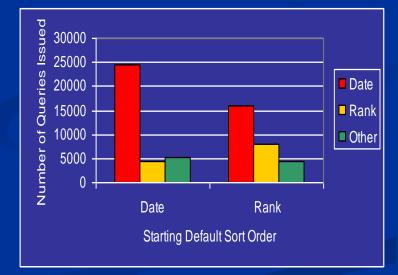
#### Top vs. Side Views

Feb       Test       Cost       Mary matching       Clear All         3038 rows refurned       Image: Cost All of the Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         11       Image: Cost All of the Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2010       Stuff I Ves Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2011       Stuff I Ves Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2012       Stuff I Ves Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2013       Image: Cost All of the Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2014       Image: Cost All of the Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2014       Image: Cost All of the Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2015       Image: Cost All of the Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2016       Image: Cost All of the Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2017       Image: Cost All of the Secon       Image: Cost All of the Secon       Image: Cost All of the Secon         2017       Ima							C	f "ve Seen	<mark>≫</mark> S*uff
Image: Stuff Ive S en       Image: Stuff Ive S en         Image: Stuff Ive S en       Image: Stuff Ive S en <th></th> <th></th> <th>Ves</th> <th></th> <th></th> <th></th> <th></th> <th>w chemic de la Coloria</th> <th></th>			Ves					w chemic de la Coloria	
3033 rows returned         If the View Opt rsb. 100         If the View Opt					Clear All	i uzzy materi 🗖			sis
Were Optings       Were Optings       Were Optings       Were Optings       Mediation       Constraints       Constraints <thconstraints< th=""> <thconstraints< th=""></thconstraints<></thconstraints<>							_	ows returned	3038 ro
<ul> <li>is</li> &lt;</ul>					Don				
3038 rows returned         Image: Stuff Ive Seen         Im					NGI	Jale vs.	DVL	e View Options., Helr	File
Stuff Ive Seen       Image: Stuff Ive Seen         Image: Stuff Ive Seen       Image: Stuff Ive Seen     <		dexing (8941)	Index	)	tch 💌 📿 Clear All	💌 🜔 Go 🛛 Fuzzy Mat			
Be yew Tools Help         Importation overload								38 rows returned	± 🗹 30
Important	_ 🗆 🔼							🔑 Stuff I've Seen	
V       287 rows returned       Stuff I've Seen         Document       Date       Ele yew Tools Help         V(Al) (287)       V(Al) (287)       V(Al) (287)         V(Al) (287)       V(Al) (287)       V(Al) (287)         V       Provide Pages (9)       Today (0)       Vesterday (0)         V       Provide Pages (9)       Today (0)       Vesterday (0)       Vesterday (0)         V       Provide Pages (9)       Older than 30 day       Pocument       Date       Rank       Path       Author         VMI 1900       VMI 190								<u>F</u> ile <u>V</u> iew Tools <u>H</u> elp	
<ul> <li>287 rows returned</li> <li>287 rows returned</li> <li>287 rows returned</li> <li>287 rows returned</li> <li>280 rows returned</li> <li>290 rows returned</li> <li>290 rows returned</li> <li>200 rows ret</li></ul>				ar All	ect Match 👻 🕻 Clea	Go Exa		information overload	<b>#</b> +
Image: Construction	_ 🗆 🔀					Stuff I've Seen		287 rows returned	+
Implementation       Implementation       Implementation       Implementation       Implementation         Implementation       Implementation       Implementation       Implementation       Implementation         Implementation       Implementation       Implementation       Implementation       Implementation         Implementation       Implementation       Implementation       Implementation       Implementation       Implementation         Implementation       Implementation       Implementation       Implementation       Implementation       Implementation       Implementation         Implementation						<u>File View Tools H</u> elp		Document	đ
Image: Search Day #546 - To Goode, a.m. 6/9/2003 10:43 AM       Image: Search Day #546 - To Goode, a.m. 6/9/2003 10:48 PM       Image: Search Day #546 - To Goode, a.m. 6/9/			▼ Go Exact Match ▼ (Clear All)			information overload			
Image: Search Day #346 - To Goose, an 6//22003 148 PM       6//2003 148 PM       328//2003 10:23 AI       395       my oncuments/sis/dataformu-032103         Image: Search Day #346 - To Goose, an 6//22003 148 PM       6//2003 148 PM       328//2003 10:23 AI       395       my oncuments/sis/dataformu-032103         Image: Search Day #346 - To Goose, an 6//22003 148 PM       6//22003 10:33 AI       395       my oncuments/sis/dataformu-032103         Image: Search Day #346 - To Goose, an 6//22003 10:33 AI       90       10//22/2003 10:23 AI       90       90//22/2003 10:23 AI         Image: Search Day #346 - To Goose, an 6//22003 10:33 AI       90       90//22/2003 10:25 AI       193       pers/al folders/ir/ms nisd//hul/relevance         Image: Search Day #346 - To Goose, an 6//22003 148 PM       6//22003 148 PM       328//2003 10:25 AI       193       pers/al folders/ir/ms nisd//hul/relevance         Image: Search Day #346 - To Goose, an 6//22003 128 PM       90       10//22/20//2003 148 PM       10//22/20//2003 148 PM       10//22/20//2003 148 PM       10//22/2003 12//2003 149 PM       10//22/20//2003 149 PM       10//22/2003 12//2003 149 PM       10//22/2003 149 PM       10//22/						190 rows returned			Ľ
<ul> <li>Image: Construction of the second s</li></ul>	Mail To	Author	nk Path	🔊 R	Date	Document		+ 🖌 Files <b>(96)</b>	£
Web Pages (9)       Today (0)         Web Pages (9)       Today (0)         Web Pages (9)       Today (0)         Web Pages (9)       Yesterday (0)         Yesterday (0)       Yesterday (0)         Yesterday (0)       Yesterday (0)         Yesterday (1)			-						🖂 a
Image: Net: FW: HL1-NAACL Program Co7/3/2003 12:27 PM       + Image: FW: HL1-NAACL Program Co7/3/2003 10:56 AM         Image: FW: HL1-NAACL Program Question       6/25/2003 10:33 AN         Image: FW: HL1-NAACL Program Question       6/25/2003 10:32 AN         Image: FW: HL1-NAACL Program Question       6/2/2003 10:32 AN         Image: FW: HL1-NAACL Program Question       6/2/2003 10:32 AN         Image: FW: HL1-NAACL Program Question       6/2/2003 10:32 AN         Image: FW: FW: HL1-NAACL Program Question <td< th=""><th></th><th></th><th></th><th></th><th></th><th>Web Pages (9)</th><th></th><th></th><th>5</th></td<>						Web Pages (9)			5
<ul> <li>He: FW: FLI-IVAALL Program Co//3/2003 10:36 AM</li> <li>He: FW: FLI-IVAALL Program Question</li> <li>6/25/2003 10:33 AN</li> <li>Older than 30 day</li>     &lt;</ul>							7/3/2003 12:27 PM	RE: FW: HLT-NAACL Program C	
<ul> <li>Re: HLT Program Question</li> <li>Maxe: HLT Program Question</li></ul>					Last 30 days (0)		.7/3/2003 10:56 AM	Re: FW: HLT-NAACL Program Co.	
<ul> <li>Javed Aslam: Publications</li> <li>6/17/2003 2:51 PM</li> <li>Jogs-sis queryissued_all bt</li> <li>Jogs-sis queryissued_all bt</li></ul>									
Summary of Cochlea 6/10/2003 4:03 PM Search Day #546 - To Googe, an 6/9/2003 10:43 AM gridviz_ozchi2003_v1 6/6/2003 1:43 AM gridviz_ozchi2003_v1 6/6/2003 1:48 PM gridviz_ozchi2003_v1 6/2/2003 2:33 PM gridviz_gridvidvidvidvidvidvidvidvidvidvidvidvidvi	^							Re: HLT Program Question	
Search Day #546 - To Google, an 6/9/2003 10:43 AM pridviz_ozchi2003_v1 6/6/2003 1:48 PM pridviz_ozchi2003_v1 6/6/2003 1:48 PM pridviz_ozchi2003_v1 6/6/2003 1:48 PM pridviz_ozchi2003_v1 6/2/2003 2:33 PM prime and folders/ir/keepingfoundthingsfound prime and folders/ir/ms stuffiveseen prime and folders/ir/ms stuffive									
Image: Seattle Times Interview /E 5/23/2003 4:35 PM       6/6/2003 1:48 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/10/2000 1:32 P         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2000 5:34 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2000 5:34 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2000 5:34 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 4:35 PM       7/11/2003 5:16 PM         Image: Seattle Times Interview /E 5/23/2003 8:57 AW       7/11/2003 5:16 PM       5         Image: Seattle Times Interview /E 5/23/2003 8:57 AW       7/11/2003 8:57 AW       53       personal folders/ir/ms stuffiveseen         Image: Seattle Times Inter							. 6/9/2003 10:43 AM	SearchDay #546 - To Googe, an	
Image: Seattle Times Interview/E 5/23/2003 4:35 PM       Image: Seattle Times Interview/E 5/23/2003 5:16 PM       Image: Seattle Times Interview/E 5/					3/28/2003 10:02 A 1	🗐 queries.txt			1 1
Image: Non-Wedge interchange - executiv									
Recap: Seattle Times Interview/E b/23/2003 4:35 PM       2/8/2002 2:57 PM       53       pe sonal folders/ir/ms stuffiveseen         Image: Seattle Times Interview/E b/23/2003 4:35 PM       2/8/2002 2:57 PM       53       pe sonal folders/ir/ms stuffiveseen         Image: Seattle Times Interview/E b/23/2003 8:57 AM       2/8/2002 1:44 PM       53       pe sonal folders/ir/ms stuffiveseen/feedback-alp				75					
Recap: Seattle Times Interview/E b/23/2003 4:35 PM       2/8/2002 2:57 PM       53       pe sonal folders/ir/ms stuffiveseen         Image: Seattle Times Interview/E b/23/2003 4:35 PM       2/8/2002 2:57 PM       53       pe sonal folders/ir/ms stuffiveseen         Image: Seattle Times Interview/E b/23/2003 8:57 AM       2/8/2002 1:44 PM       53       pe sonal folders/ir/ms stuffiveseen/feedback-alp				64					Ľ
Recap: Seattle Times interview/E b/23/2003 4:35 PM 2nd brain 2/8/2002 2:57 PM 53 pe sonal folders/ir/ms stuffiveseen 19/2003 8:57 AM 2nd brain 8/26/2002 1:44 PM 53 p. isonal folders/ir/ms stuffiveseen/feedback-alp				64					
Recap: Seattle Times Interview/E b/23/2003 4:35 PM       2/8/2002 2:57 PM       53       pe sonal folders/ir/ms stuffiveseen         Pasist03       19/2003 8:57 AW       2/8/2002 1:44 PM       53       pe sonal folders/ir/ms stuffiveseen/feedback-alp			my ocuments \arrenceengine \fileswebqueries	60				Coverage: Seattle Times (Memory)	
				53			5/23/2003 4:35 PM	Recap: Seattle Times Interview/E	
				53			V19/2003 8:57 AM	颵 jasist 03	
Electosure b - revised in response 7/7/1999 1:09 PM 52 revisional folders/nrc-diggovt			rersonal folders/nrc-diggovt	52				<	
Grant 16 9/9/2000 3:38 PM 52 personal folders/ir/keepingfoundthingsfound				52		igrant 16			
grant 17-finalsubmitted 12/18/2000 9:57 AM 45 my documents papers keepingfoundthingsfound William Jones		Nilliam Jones		45	12/18/2000 9:57 AM	grant 17-finalsubmitted			
grant 17-finalsubmitted       12/18/2000 9:57 AM       45       my documents\papers\keepingfoundthingsfound       William Jones         grant 17       12/18/2000 10:58 AM       45       personal folders/ir/keepingfoundthingsfound				45	12/18/2000 10:58 AM	🕎 grant 17			
	>								

## SIS Usage Data, cont'd

#### <u>UI Usage</u>

- 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

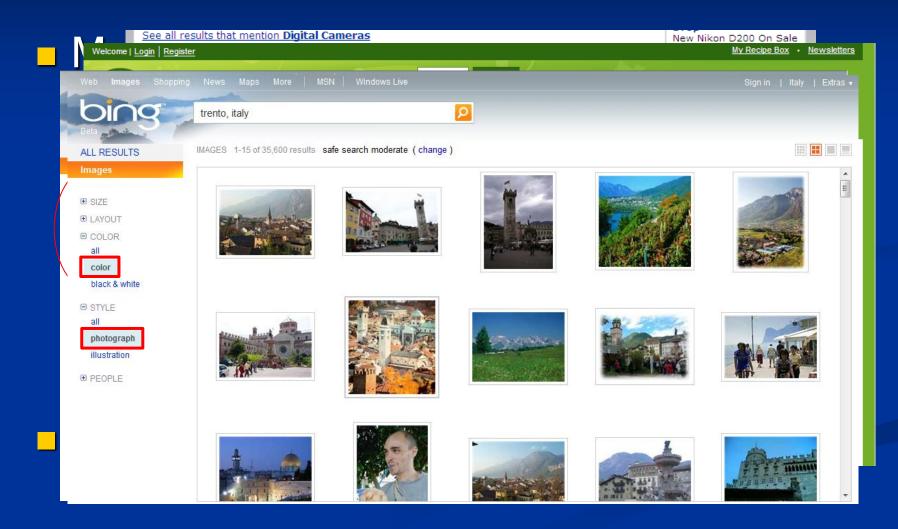
- "Useful date" is dependent on the object !
  - Appointment, when it happens
  - File, when it is changed
  - Email and Web, when it is seen
- "People" attribute vs. contains
  - To, From, Cc, Author, Artist

## Ranked list vs. Metadata (for personal content ... and beyond)

Google Desktop Search results	
Google Web Images Groups News Froogle Local Desktop more a	
Desktop Stuff I've Seen	
Desktop: All - 7 emails - 8 files - 22 web	
Going on vacation	
Why Rich Metadata?	Go
<ul> <li>People remember many attributes i</li> </ul>	n re-finding
- Often: time, people, file type, etc.	
Seldom: only general overall topi	C
<ul> <li>Rich client-side interface</li> </ul>	
- Support fast iteration/refinement	
<ul> <li>Fast filter-sort-scroll vs. next-nex</li> </ul>	t-next
***       p212fig4.jpg       10/10/2005       12.5 KB         Comments:       Keywords:       Keywords:       Nu Documents/papers/Glaccows-TBY/SIGTP 2004Workshop-And-Others/Information ceeking recearch peeds ex	a friend in the hundreds of gigabytes of storage now common on personal computers can be an exercise in patience

Teevan et al., HCIR 2008

## Metadata on the Web



Teevan et al., SIGIR 2007

## **Re-finding on the Web**

50-80% page visits are re-visits
 30-50% of <u>queries</u> are re-finding queries

		Repeat Click	New Click
Repeat Query	33%	29%	4%
New Query	67%	10%	57%
		39%	61%

Total = 43% Big opportunity to support re-finding on the Web

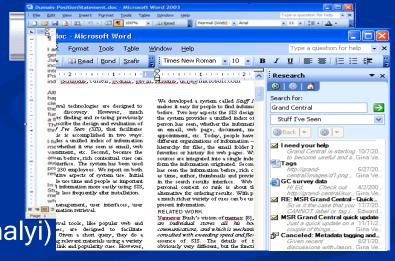
## Search in Task Contexts

#### 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



## Documents as (a simple) Context

Proactive "query" specification depending on current document content and activities

> File G

#### Recommendations

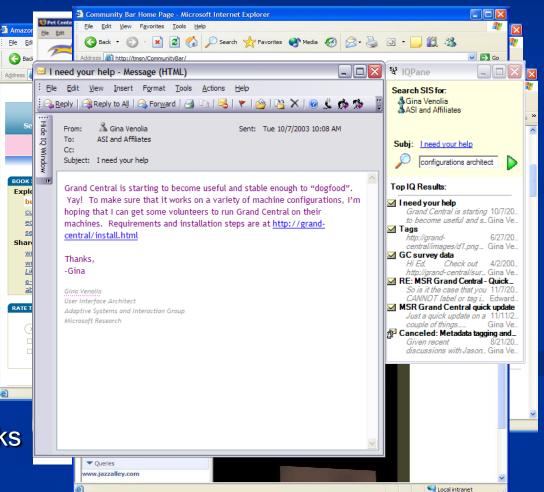
People who bought this also bought ...

#### **Contextual Ads**

Ads relevant to page

#### **Community Bar**

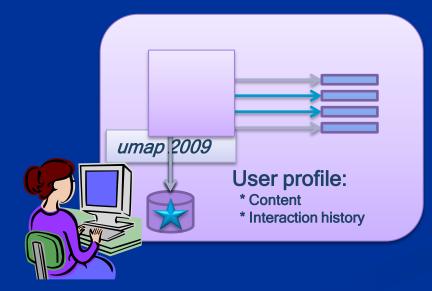
- Context search, Notes, Chat, Tags, Inlinks, Queries
- http://www.communitybar.net
- Implict Queries (IQ)
  - Also Y!Q, Rememberance Agent, Watson, Query-free search
- Even more possibilities for context-driven retrieval w/ rich sensors and ubiquitous networks



Teevan et al., SIGIR 2005

### PSearch: Personalized Search (Much Richer Context)

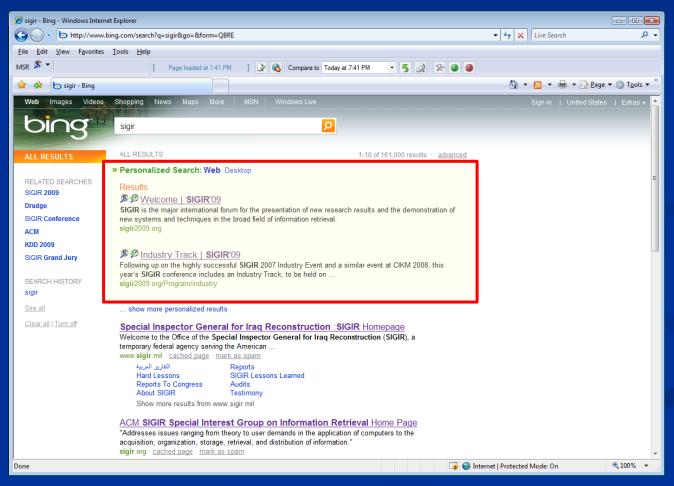
- <u>Today</u>: People get the same results, independent of current session, previous search history, etc.
- PSearch: Uses rich client-side model of a user to personalize search results





## **PSearch Demo**

#### Query: *SIGIR*



## PSearch ... How it Works

Key components and challenges Building a user profile Personalizing the ranking Personalizing the presentation When to personalize? Beyond PSearch ... Other types of profiles and applications Privacy and security Evaluation End of serendipity?

## **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
  - Server: Richer communities, portability

#### **PSearch**

## **Personalized Ranking**

4 .	
sigir	<b>Q</b>
ALL RESULTS	1-10 of 185,000 results
	eral for Iraq Reconstruction SIGIR Homepage pecial Inspector General for Iraq Reconstruction (SIGIR), a ng the American
التقارير الحربية Hard Lessons Reports To Congress About SIGIR	Reports SIGIR Lessons Learned Audits Testimony
Show more results from w	www.sigir.mil

ACM SIGIR Special Interest Group on Information Retrieval Home Page

"Addresses issues ranging from theory to user demands in the application of computers to the acquisition, organization, storage, retrieval, and distribution of information." sigir.org - cached page

#### SIGIR'08 - Singapore

SIGIR is the major international forum for the presentation of new research results and for the demonstration of new systems and techniques in the broad field of information ... www.sigir2008.org - cached page

#### Welcome | SIGIR'09

SIGIR is the major international forum for the presentation of new research results and the demonstration of new systems and techniques in the broad field of information retrieval. sigir2009.org · <u>cached page</u>

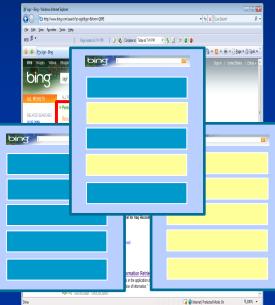
#### SIGIR 2004

The major international forum for the presentation of new research results and demonstration of new systems and techniques in information retrieval. The 27th Annual International ... www.sigir.org/sigir2004 · cached page  Personal Rank = f(Cont, InterHist, Web)
 P\_Content Match: sim(result, user\_content\_profile)
 P\_InteractHist Match: visited URLs and sites
 Web Match:

web rank

## **Personalized Presentations**

# Presentation options Inline display (for demo, use) Also tried: just personalized, tabs, slider, fisheye views, metadata, ... Interleaved results (for evaluation)

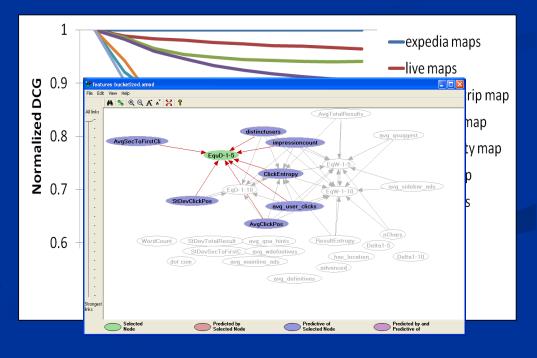


 Transparency and control
 People try to explain what they got
 Highlight personally relevant terms and sites
 ACM SIGIR Special Interest Group on Information Retrieval Home Page

"Addresses issues ranging from theory to **user** demands in the application of computers to the acquisition, organization, storage, retrieval, and distribution of information." sigir.org

## When to Personalize?

- Personalization works well for some queries,
   ... but not for others
- Framework for understanding when to personalize
- Models for predicting when to personalize (using features of <u>query</u>, <u>user-query</u>)
  - Personal ranking
    - Personal relevance (explicit or implicit)
  - Group ranking
    - Decreases as you add more people
  - Gap is "potential for personalization (p4p)"



## When to Personalize?

High impact / low risk opportunities
 "Re-finding" queries - depend on <u>user-query</u>



\* 43% coverage\* high accuracy

Personalization/presentation strategies that do not affect ranking

 Annotate results that are personally relevant (allows for filtering/sorting)

Personalize snippets to aid relevance assessment

## Beyond PSearch: Other User Profiles



PSearch

- Type of information
  - Content: Past queries, web pages, desktop
  - Behavior: Visited pages, explicit feedback Groupization
- Time frame: Short term, long term
- Who: Individual, group
- Where the profile resides:
  - Local: Richer profile, improved privacy
  - Server: Richer communities, portability

Session Memory

Query Suggest

uRank

## **Personalization and Privacy**

#### PSearch



- Local profile, local computation
- Nothing sent to the server except the original query
- Need profile and web content in same place to rank
- When information is stored in the cloud
  - Send query and user profile, or store user profile in cloud
  - Transparency
  - Control
- Other approaches we are exploring
  - Matching an individual to a group
  - Light weight profiles (e.g., queries in a session)
  - Public or semi-public profiles (e.g., Tweets, Facebook status, blogs)

## **Evaluating Personalized Search Systems**

to://search.msn.com/results.asox?srch=105&EORM=AS5&

The 14th International World Wide Web Conference 2005

Contents Menu, Top, Conference Organizers ... General questions about WWW20

Contents Menu. Top Papers Call For Paper. WWW2005 CALL FOR PAPERS ... W better known as Makuhari Messe). The ... IW3C2 LIAISON TO WWW2005. Vincent

WebEngineering.org Community - The Site about Web Enginee

Sites ICWE2005 WWW2005 Get Involved Eurther information is available at

XML Conference Calendar. This is a calendar of XML related trade shows, confere

Clara California. WWW2005, Chiba, Japan, May 10-14, 2005 ... Web Services Desi

msn<sup>1</sup> Search

WWW2005 Call For Pape

www2005 org/naners/cfn.htm

XML Conference Calenda

www.ibiblio.org/xml/tradeshows.btm

Results 1-50 of about 72 containing "www2005

www2005

- Explicit judgments (offline and *in situ*)
  - Evaluate components before system
  - NOTE: What's relevant MSN Search: www2005 -- More Useful Everyday - Microsoft Internet Explorer File Edit View Favorites Tools Help
- Deploy system

UMAP, June 24 2009

- Verbatim feedback.
- Measure behaviora reformulation, aban Click biases - order, Interleaving for unbit Link implicit and exp Beyond a single qu

Elier Tutorials ( WWW2005 International WWW Conference Committee Workshops as Makuhari I Panels special intere ... 2004 - New York (US) WWW2005. 2006: Edinburgh workshops Sort of Curious Browser Study (~4k) Call to Host a Conference the next Conferences is: V \* 45% w/ just click \* 75% w/ click + dwell + session

www

2005

Call For Pape

Paper Tracks

Track Chairs

Information

Submission

Submission Site

Ton

Papers

🕜 🔎 Search 🤺 Favorites 🚱 🔗 - 🎽 🗿 - 🗾 🏭 🍇 🥹

..

E Go

Top 
 Papers Call For Pap

😰 🏠 🔎 Search 🤺 Favorites 🚱 😒 - 🍓 💿 🔹 🔜 🎇 🦓 🥶

Curious Browser

You searched for: www2005

Search result: WWW2005 Call For Paper

result?

WWW200

The Internation

cordially invite

Conference o

The first intern

born Since th

industries to p

Web

The 14th International World Wide Web Conference

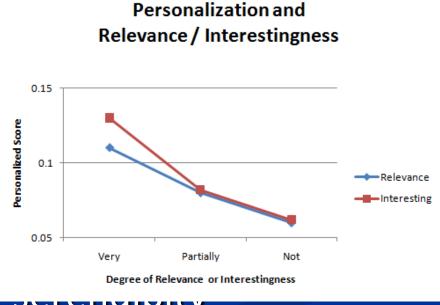
Did you find the information you needed at this search

## End of Serendipity?



Does great search and personalization mean the end of serendipity?

No, actually better point
 Relevance vs. interesti
 Personalization finds m
 Personalization finds m
 Many not relevant resu
 Need to be "ready" for



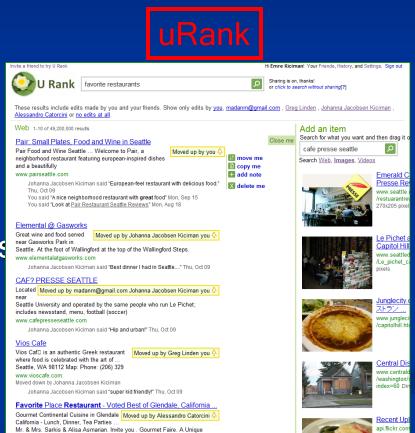
## Beyond Search - Sharing & Collaborating

uRank - Edit, organize and share search results w/ your friends

Edit: move results up/down, annotate w/ notes

Results as first class objects

- Lists: create lists while you're researching
- Sharing: share results, lists w/ your friends



Try It: http://research.microsoft.com/projects/urank

## 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



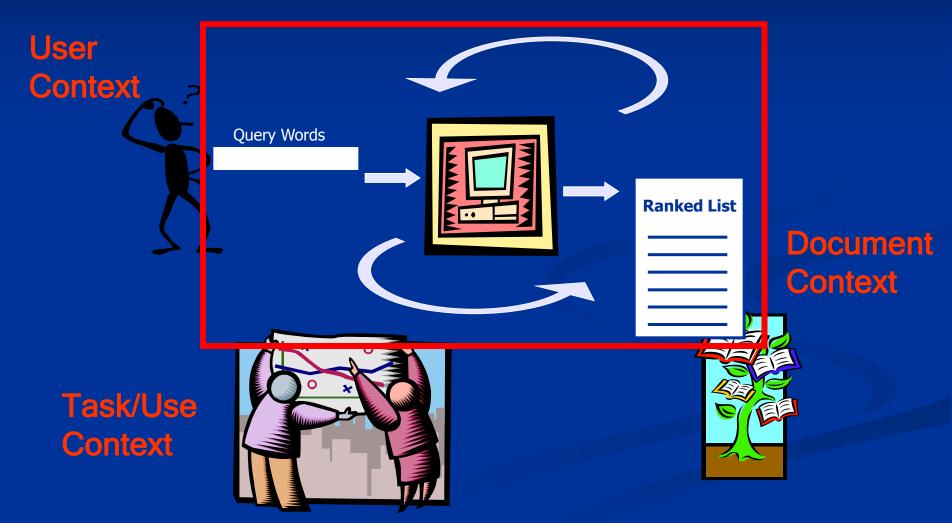
Figure 1. The SearchTogether client. (a) integrating messaging, (b) query awareness, (c) current results, (d) recommendation queue, (e)(f)(g) search buttons, (h) page-specific metadata, (i) toolbar, (j) browser

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

## Looking Ahead ...

- Continued advances in scale of systems, diversity of resources, and quality of ranking, etc.
- Tremendous <u>new opportunities</u> to support information retrieval and analysis 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 analysis, use and sharing of results
  - Considering search as part of richer landscape

## Thinking Outside the (Search) Box



## Thank You !

Questions/Comments ...

More info, <u>http://research.microsoft.com/~sdumais</u>

#### Stuff I've Seen & Memory Landmarks

- S. Dumais, E. Cutrell, J. J. Cadiz, G. Jancke, R. Sarin & D. C. Robbins (2003). <u>Stuff I've Seen: A system for personal information</u> <u>retrieval and re-use</u>. *SIGIR 2003*.
- M. Ringel, E. Cutrell, S. Dumais & E. Horvitz (2003). <u>Milestones in time: The value of landmarks in retrieving information from personal stores.</u> *Interact 2003*.
- E. Horvitz, S. Dumais & P. Koch (2004). Learning predictive models of human memory landmarks. Cognitive Science 2004.
- Personalized Search, Groupization & Serendipity
  - J. Teevan, S. Dumais & E. Horvitz (2005). <u>Personalizing search via</u> <u>automated analysis of interests and activities</u>. *SIGIR 2005*.
  - F. Radlinksi & S. Dumais (2006). <u>Improving personalized web</u> search using results diversification. SIGIR 2006.
  - J. Teevan, S. Dumais & E. Horvitz (2006). <u>To personalize or not:</u> <u>Modeling queries with variation in user intent</u>. *SIGIR 2008*.
  - J. Teevan, S. Dumais & E. Horvitz (in press). <u>Potential for</u> personalization. *TOCHI, in press*.
  - P. Andre, J. Teevan & S. Dumais (2009). From x-rays to silly putty via Uranus: Serendipity and its role in Web search. CHI 2009.

- J. Teevan, M. Ringel & S. Bush(2009). <u>Discovering and using groups</u> to improve personalized search. WSDM 2009.
- R. White, S. Dumais & J. Teevan (2009). <u>Characterizing the influence of domain expertise on Web search behavior</u>. *WSDM 2009.*
- Re-Fnding & Metadata on the Web
  - J. Teevan, E. Adar, R. Jones & M. Potts (2007). Information retrieval re-retrieveal: Repeat queries in Yahoo's logs. SIGIR 2007.
  - J. Teevan, S. Dumais & Z. Gutt (2008). <u>Challenges in supporting</u> <u>faceted search in large, heterogeneous corpora like the Web.</u> *HCIR.*

Implicit Queries & NewsJunkie

- S. Dumais., E. Cutrell, R. Sarin & E. Horvitz (2004). Implicit queries (IQ) for contextualized search. SIGIR 2004.
- E. Gabrilovich, S. Dumais & E. Horvitz (2004). <u>NewsJunkie: Providing</u> personalized newsfeed analysis via information novelty. WWW 2004.

#### Curious Browser

- S. Fox, K. Jamawat, M. Mydland, S. Dumais & T. White (2005). <u>Evaluating implicit measures to improve the research experience.</u> ACM:TOIS, 23(2), 2005.
- Search Together and uRank
  - M. Morris & E. Horvitz (2007). <u>SearchTogether: An interface for</u> <u>collaborative Web search.</u> UIST 2007.