

Temporal Dynamics and Information Systems

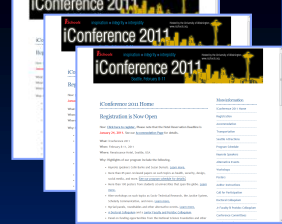
Susan Dumais

Microsoft Research

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

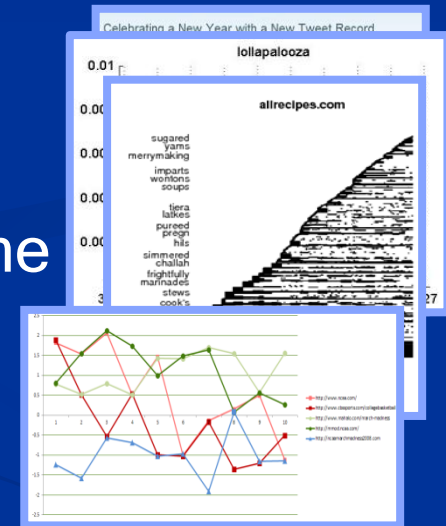
In collaboration with:

Eric Horvitz, Jaime Teevan, Eytan Adar, Jon Elsas, Ed Cutrell,
Dan Liebling, Richard Hughes, Merrie Ringel Morris, Evgeniy
Gabrilovich, Krysta Svore, Anagha Kulkani



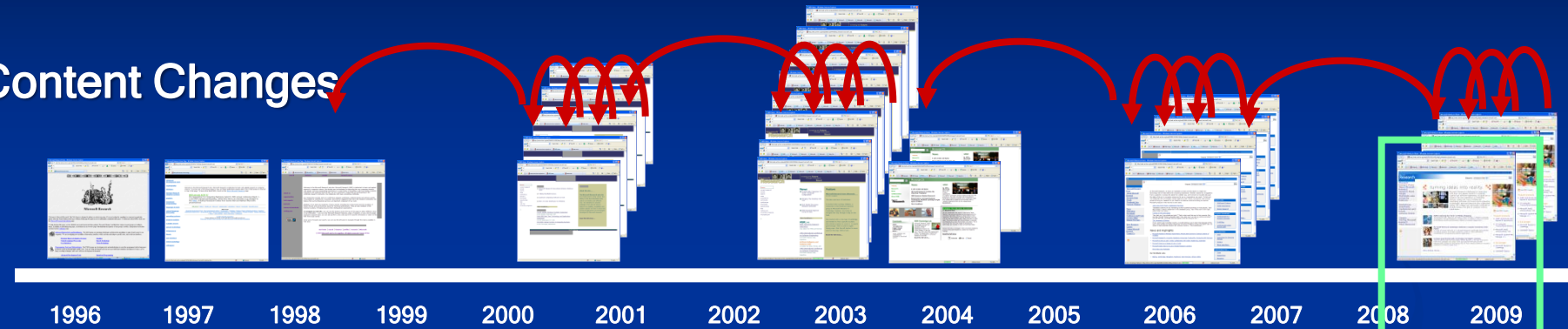
Information Dynamics

- Many differences between physical & digital libraries
- **Change is everywhere in digital information systems**
 - New documents (and queries) appear all the time
 - Query volume changes over time
 - Document content changes over time
 - What's relevant to a query changes over time
 - E.g., *U.S. Open 2010* (in May vs. Sept)
 - E.g., *Hurricane Earl* (in Sept 2010 vs. before/after)
 - User interaction changes over time
 - E.g., tags, anchor text, social networks, query-click streams, etc.
- Change is pervasive in digital information systems
... yet, we're not doing much about it !



Information Dynamics

Content Changes



1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

User Visitation/ReVisitation

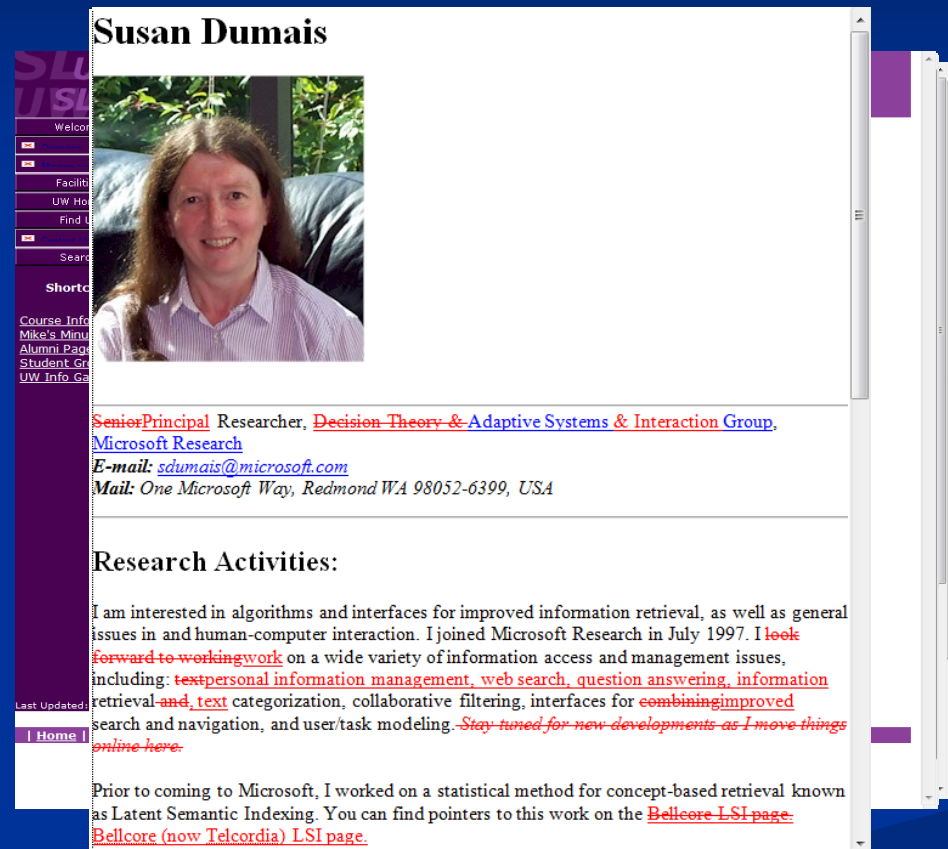
Today's Browse and Search Experiences

But, ignores ...


Digital Dynamics Easy to Capture

- Easy to capture

- But ... few tools support dynamics



Susan Dumais



SeniorPrincipal Researcher, [Decision Theory & Adaptive Systems & Interaction Group](#),
[Microsoft Research](#)
E-mail: sdumais@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 work](#) on a wide variety of information access and management issues, including: [textpersonal information management](#), [web search](#), [question answering](#), [information retrieval](#) [and text](#) categorization, collaborative filtering, interfaces for [combiningimproved](#) search and navigation, and user/task modeling. *Stay tuned for new developments as I move things [online 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 [Bellecore LSI page](#).
[Bellecore \(now Telcordia\) LSI page](#).

Overview

- Characterize change in digital content
 - Content changes over time
 - People re-visit and re-find over time
- Improve retrieval and understanding
 - Examples from our work on search and browser support ... but more general
 - Desktop: Stuff I've Seen; Memory Landmarks; LifeBrowser
 - News: Analysis of novelty (e.g., NewsJunkie)
 - Web: Tools for understanding change (e.g., Diff-IE)
 - Web: Retrieval models that leverage dynamics

Stuff I've Seen (SIS)

- Many silos of information

- SIS:

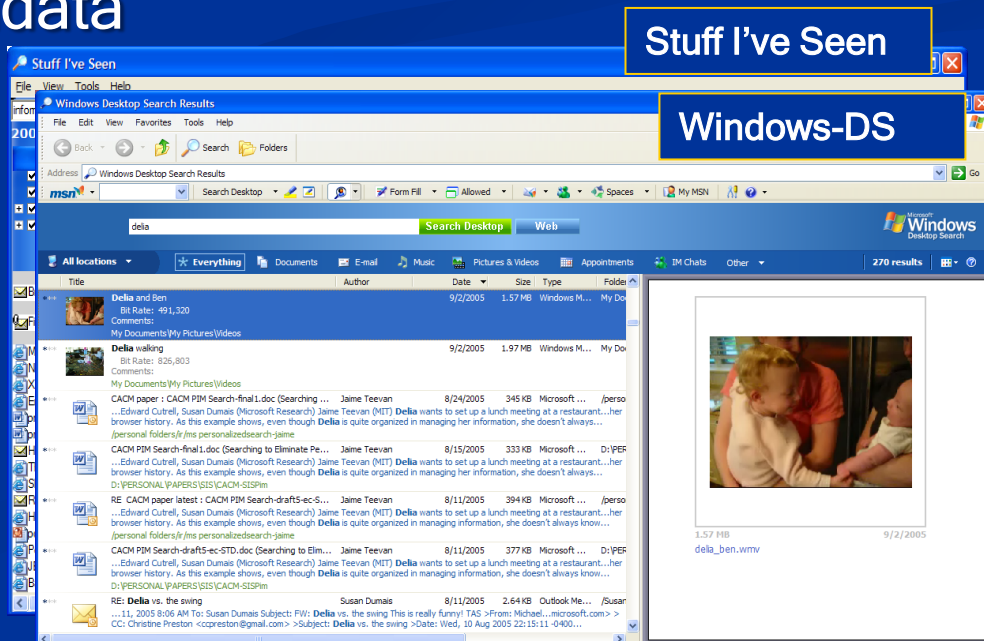
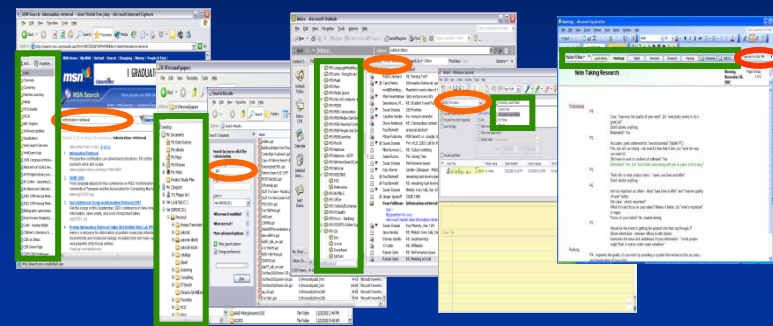
- Unified access to distributed, heterogeneous content (mail, files, web, tablet notes, rss, etc.)

- Index full content + metadata

- Fast, flexible search

- Information re-use

- SIS -> Windows Desktop Search



Example Desktop Searches

Lots of metadata
... especially time

Looking for: *recent email from Fedor that contained a link to his new demo*

Initiated from: Start menu

Query: from:Fedor

Looking for: *the pdf of a SIGIR paper on context and ranking (not sure it used those words) that someone (don't remember who) sent me about a month ago*

Initiated from: Outlook

Query: SIGIR

Looking for: *meeting invite for the last intern handoff*

Initiated from: Start menu

Query: intern handoff kind:appointment

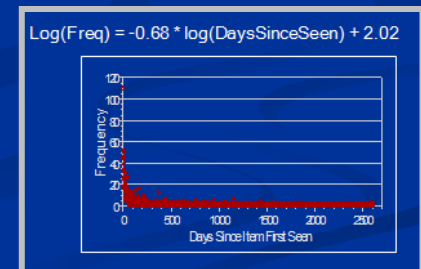
Looking for: *C# program I wrote a long time ago*

Initiated from: Explorer pane

Query: QCluster*.*

Stuff I've Seen: Findings

- Studied using: free-form feedback, questionnaires, usage patterns from log data, in situ experiments, lab studies for richer data
- Personal stores: 5k-1500k items [SD: 100k items; 1k new items/wk]
- Information needs:
 - Desktop search != Web search
 - People are important - 29% queries involve names/aliases
 - Date is the most common sort order, even w/ “best-match” default
 - Few searches for “best” matching object
 - Many other criteria (e.g., time, people, type), depending on task
 - Need to support flexible access
 - Abstractions important - “useful” date, people, pictures
 - Age of items retrieved
 - Today (5%), Last week (21%), Last month (47%)
 - Need to support episodic access to memory

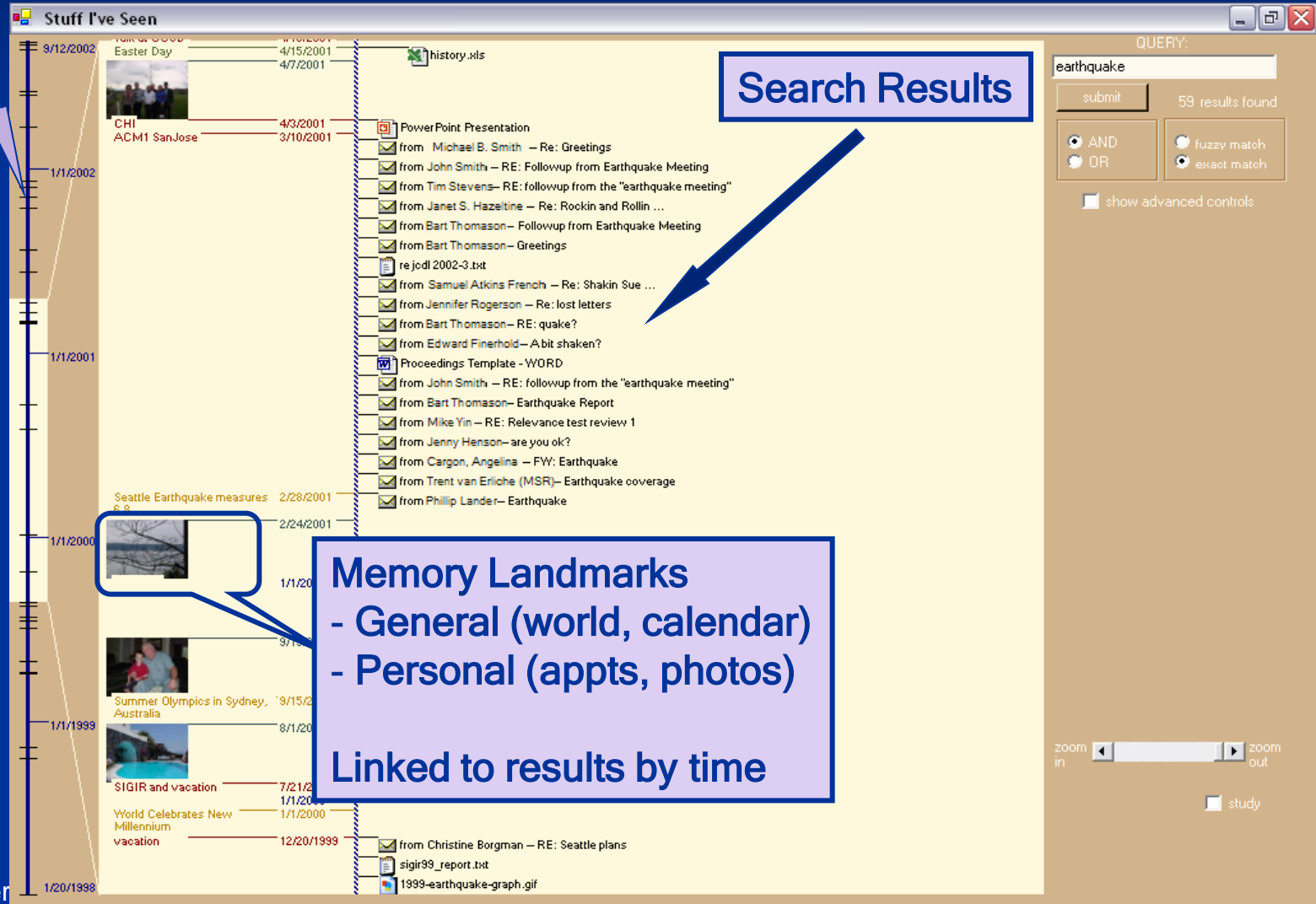


Memory Landmarks

- Importance of episodes in human memory
 - Memory organized into episodes (Tulving, 1983)
 - People-specific events as anchors (Smith et al., 1978)
 - Time of events often recalled relative to other events, historical or autobiographical (Huttenlocher & Prohaska, 1997)
- Identify and use landmarks facilitate search and information management
 - Timeline interface, augmented w/ landmarks
 - Learn Bayesian models to identify memorable events
- Extensions beyond search, e.g., Life Browser

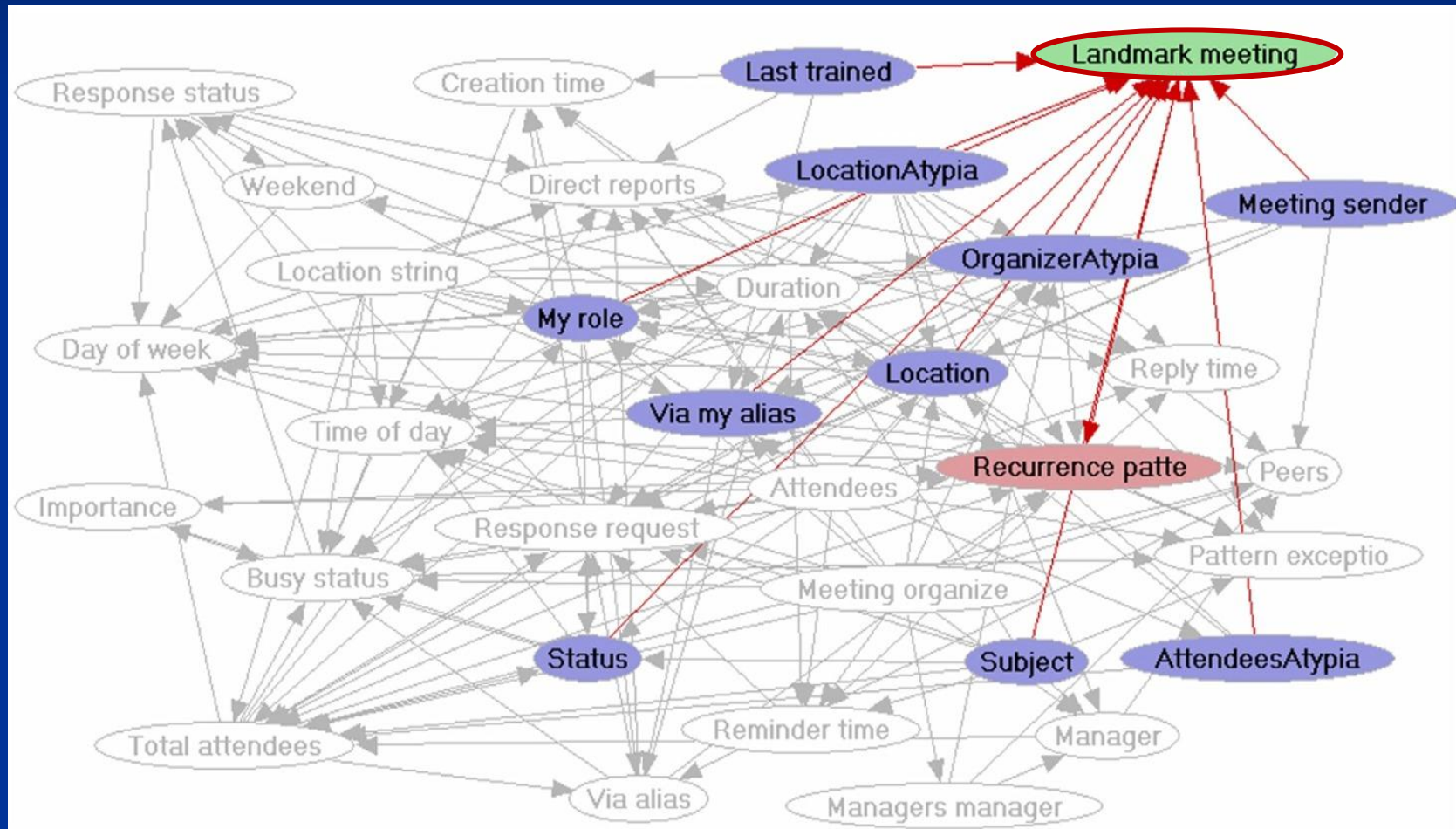
Memory Landmarks

Distribution of Results Over Time

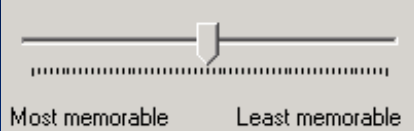


Memory Landmarks

Learned models of memorability






LifeBrowser



Most memorable Least memorable

Query

Date	Events	Results by Time
Feb 4 2005 Fri	 <p>privacy talk 2005.ppt priv_out.ppt</p>	<ul style="list-style-type: none"> Micronews 5 to 1 JamBayes Discussion and Demo JamBayes Discussion and Demo
	Micronews 5 to 1.doc	
Thu	 <p>lake-central-visualize.ppt C4 AUV.doc</p> <p>NRAC Italy Meeting</p>	<ul style="list-style-type: none"> Inrix folks 6 month check up...> Video... smartphlow.mpg
Wed	<p>isometric_volumes_study.p LC_visual.ppt techfest.doc</p>	<ul style="list-style-type: none"> Phlat c# filter & query controls Phlat c# filter & query controls Smartphlow/Caravel meeting recap
Tue	 <p>Gates Invitation.doc</p>	<ul style="list-style-type: none"> Smartphlow cannot start? Smartphlow cannot start? 6 month check up...> smartphlow_exclamation.bmp
	20040525_C10.xls	
Mon	<p>Farewell Party for David Weise</p> <p>MSMLS coordination</p>	<ul style="list-style-type: none"> Asi dev status week of 1/17/2005 Questions about SmartPhlow

Images
& videos

Desktop
& search activity

Appts &
events

Locations

Whiteboard
capture

NewsJunkie

Evolution of Context over Time

- News is a stream of information w/ evolving events
 - But, it's hard to consume it as such
 - Personalized news using information novelty
- Identify clusters of related articles
- Characterize what a user knows about an event
- Compute the novelty of new articles, relative to this background (relevant & novel)
 - Novelty = $KLDivergence(article || current_knowledge)$
- Use novelty score and user preferences to guide what, when, and how to show new information

NewsJunkie in Action

NewsJunkie: Pizza delivery man w/ bomb incident



Characterizing Web Change

Content Changes



- Large-scale Web crawls, over time
 - Revisited pages
 - 55,000 pages crawled hourly for 18+ months
 - Unique users, visits/user, time between visits
 - Pages returned by a search engine (for ~100k queries)
 - 6 million pages crawled every two days for 6 months

Measuring Web Page Change

- Summary metrics
 - Number of changes
 - Amount of change
 - Time between changes
- Change curves
 - Fixed starting point
 - Measure similarity over different time intervals
- Within-page changes

Measuring Web Page Change

■ Summary metrics

■ Number of changes

- 33% of Web pages change
- 66% of visited Web pages change
 - 63% of these change every hr.

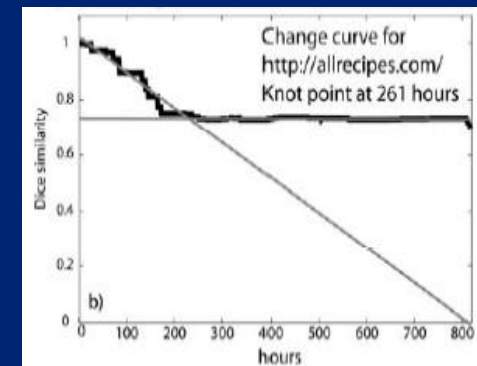
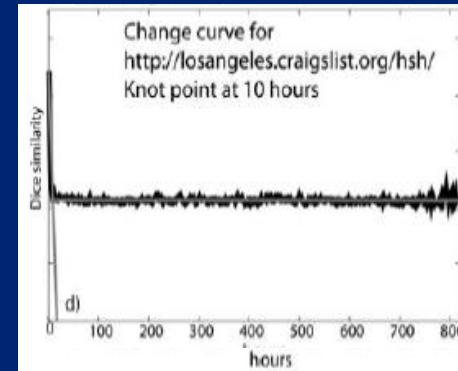
■ Amount of change

■ Time between changes

- Avg. Dice coeff. = 0.80
- Avg. time bet. change = 123 hrs.
- .edu and .gov pages change infrequently, and not by much
- popular pages change more frequently, but not by much

Measuring Web Page Change

- Summary metrics
 - Number of changes
 - Amount of change
 - Time between changes
- Change curves
 - Fixed starting point
 - Measure similarity over different time intervals



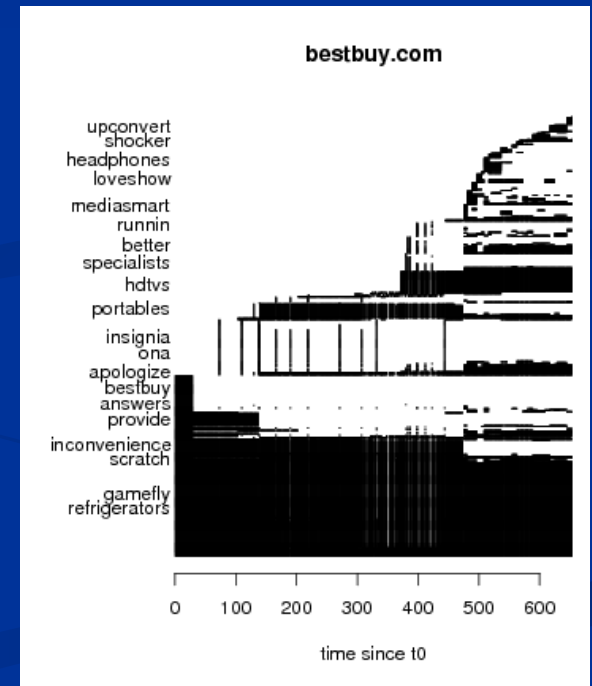
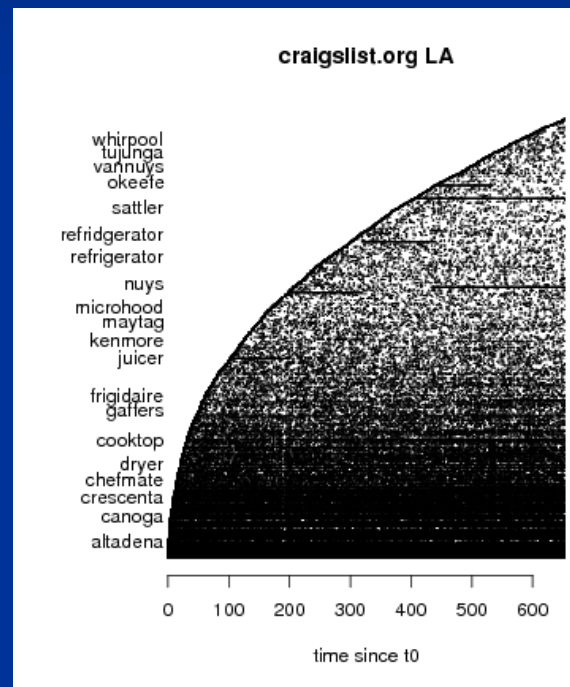
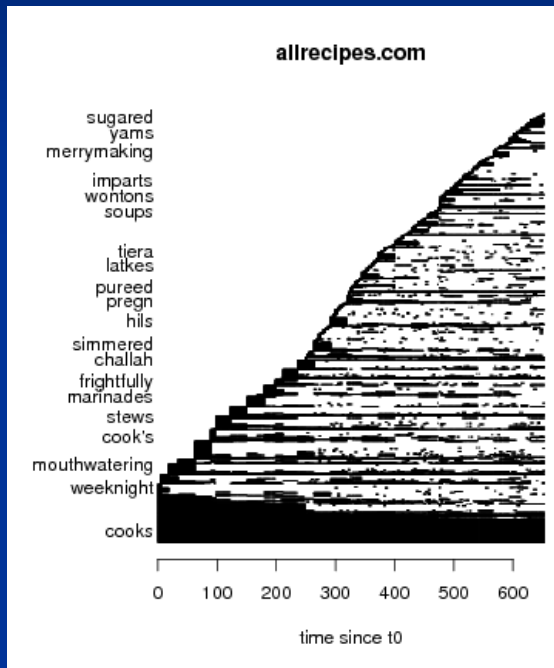
Measuring Within-Page Change

- DOM-level changes
- Term-level changes
 - Divergence from norm
 - cookbooks
 - salads
 - cheese
 - ingredient
 - bbq
 - ...
 - “Staying power” in page



Sep. Oct. Nov. Dec.
Time

Example Term Longevity Graphs



Revisitation on the Web

- Revisitation patterns
 - Log analyses
 - Toolbar logs for *revisitation*
 - Query logs for *re-finding*
 - User survey to understand intent in revisitations

1996

1997

1998

1999

2000

2001

2002

2003

2004

2005

2006

2007

2008

2009



User Visitation/ReVisitation

What was the last Web page you visited?

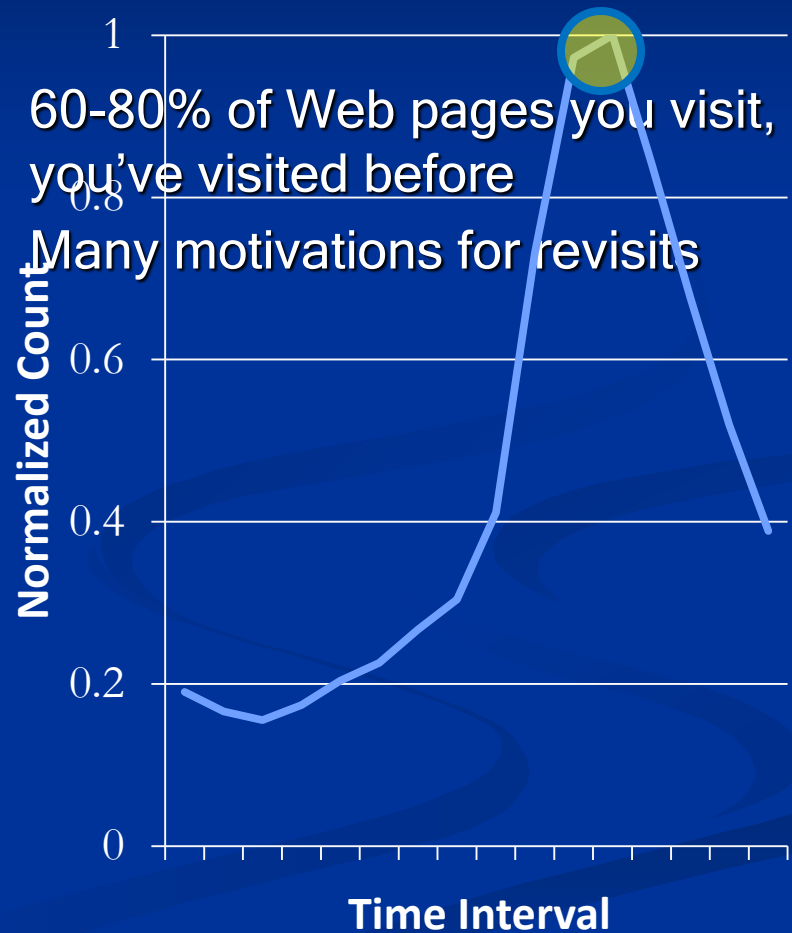
Why did you visit (re-visit) the page?

Measuring Revisitation

- Summary metrics
 - Unique visitors
 - Visits/user
 - Time between visits
- Revisitation curves
 - Histogram of revisit intervals
 - Normalized



- 60-80% of Web pages you visit, you've visited before
- Many motivations for revisits



Possible Relationships Between Change and Revisitation

iSchools inspiration • integrity • intrepidity
Hosted by the University of Washington
www.ischools.org

iConference 2011

Seattle, February 8-11

iConference 2011 Home

Registration is Now Open

How: [Click here to register](#). Please note that the Hotel Reservation Deadline is **January 24, 2011**. See our [Accommodation Page](#) for details.

What: iConference 2011
When: February 8-11, 2011
Where: Renaissance Hotel, Seattle, USA

Why: Highlights of our program include the following.

- Keynote speakers Colin Burke and Susan Dumais. [Learn more.](#)
- More than 85 peer-reviewed papers on such topics as health, security, design, social media, and more. [See our program schedule for details.](#)
- More than 100 posters from students at universities that span the globe. [Learn more.](#)
- Nine workshops on such topics as Socio-Technical Research, the Justice System, Scholarly Communication, and more. [Learn more.](#)
- Myriad panels, roundtables and other alternative events. [Learn more.](#)
- [A Doctoral Colloquium](#) and a [Junior Faculty and Postdoc Colloquium](#).
- Panel on funding opportunities from the National Science Foundation and other

More information

- [iConference 2011 Home](#)
- [Registration](#)
- [Accommodation](#)
- [Transportation](#)
- [Seattle Attractions](#)
- [Program Schedule](#)
- [Keynote Speakers](#)
- [Alternative Events](#)
- [Workshops](#)
- [Posters](#)
- [Author Instructions](#)
- [Call for Participation](#)
- [Doctoral Colloquium](#)
- [Jr Faculty & Postdoc Colloquium](#)
- [Conference Committees](#)

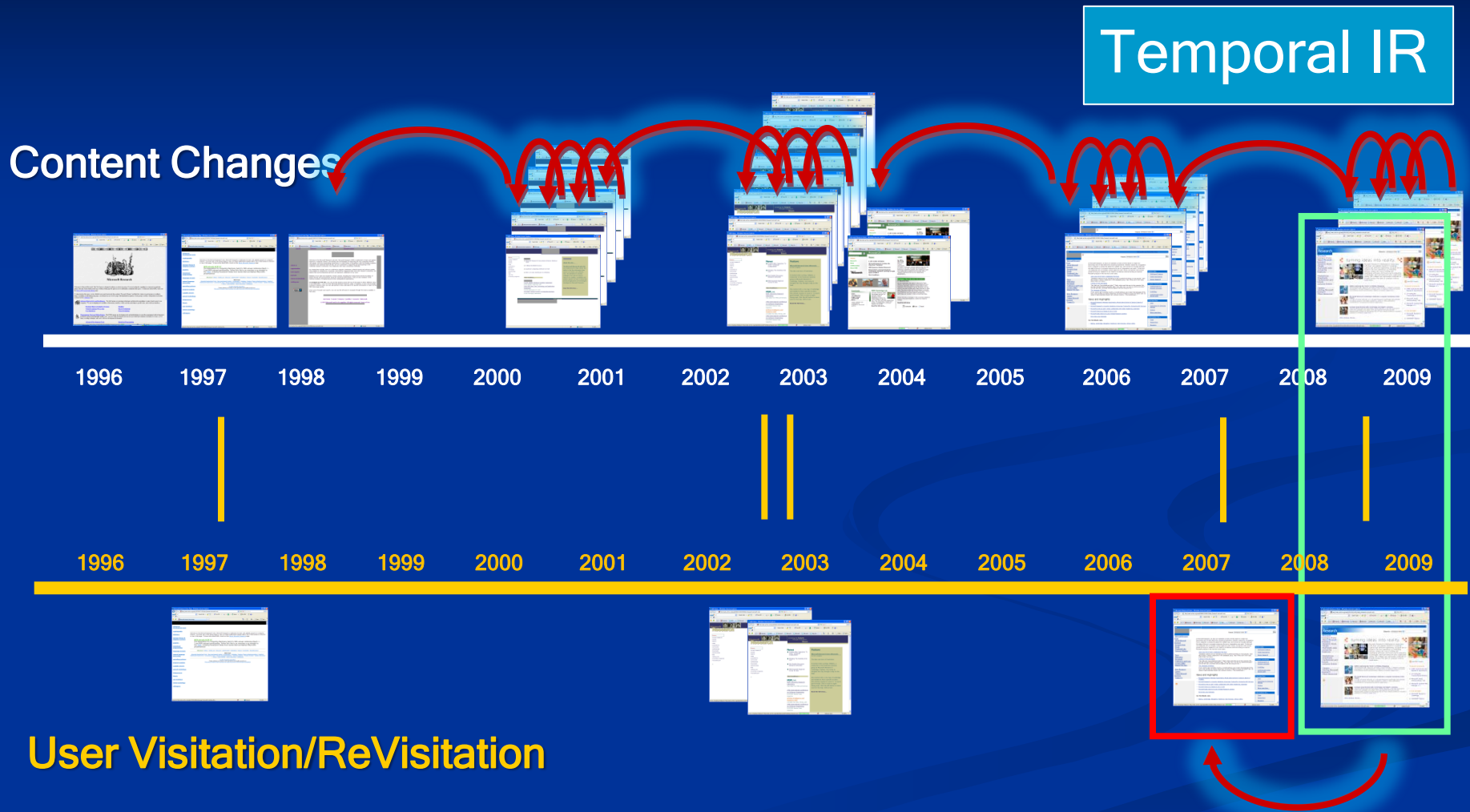
- Interested in change
 - Monitor
- Effect change
 - Transact
- Change unimportant
 - Re-find old
 - Change can interfere with re-finding

Revisitation and Search (ReFinding)

- Repeat query (33%)
 - Q: *iconference 2011*
- Repeat click (39%)
 - <http://www.ischools.org/iConference11>
 - Q: *iconference 2011; iconference*
- Big opportunity (43%)
 - 24% “navigational revisits”

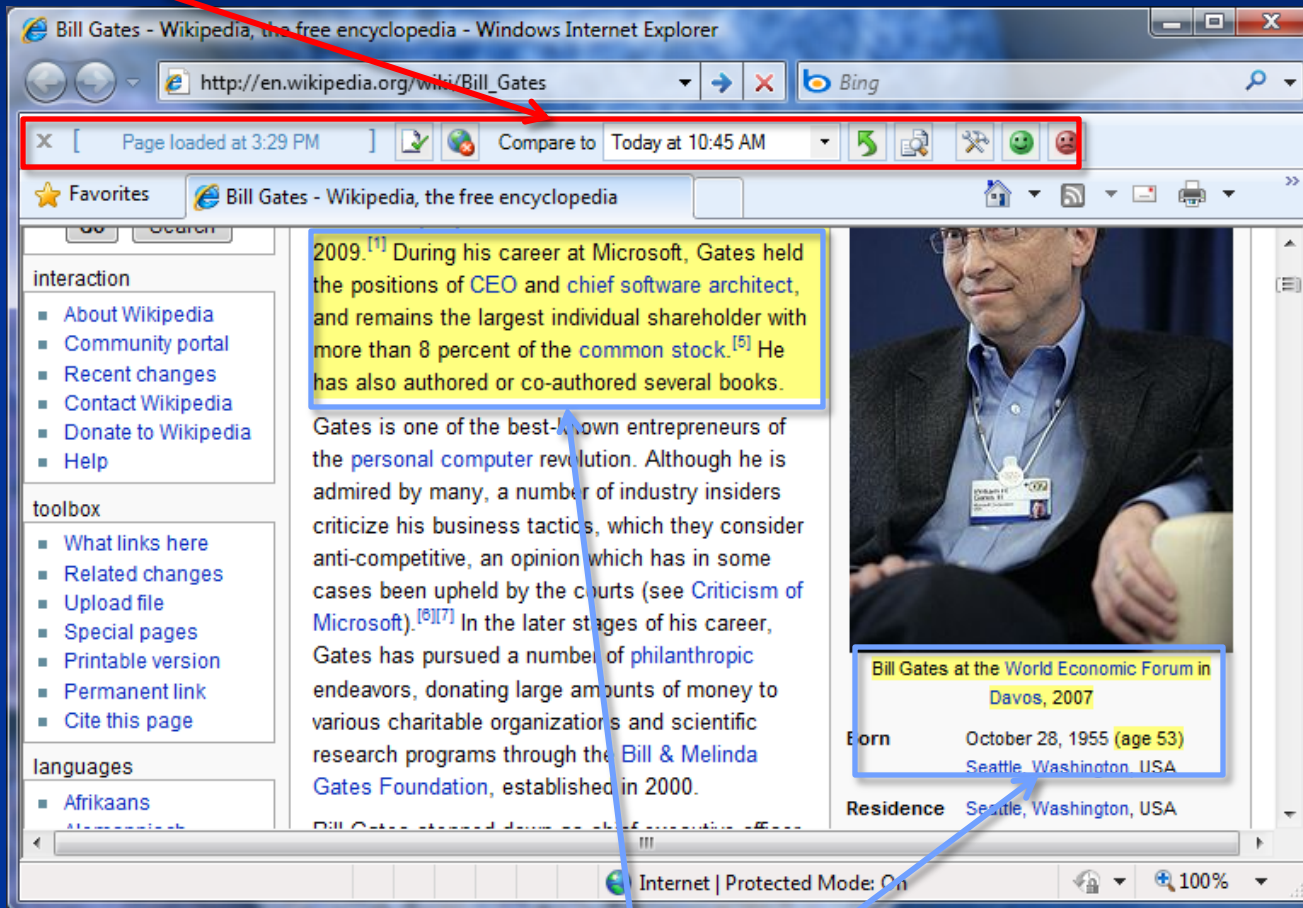
		Repeat Click	New Click
Repeat Query	33%	29%	4%
New Query	67%	10%	57%
		39%	61%

Building Support for Web Dynamics



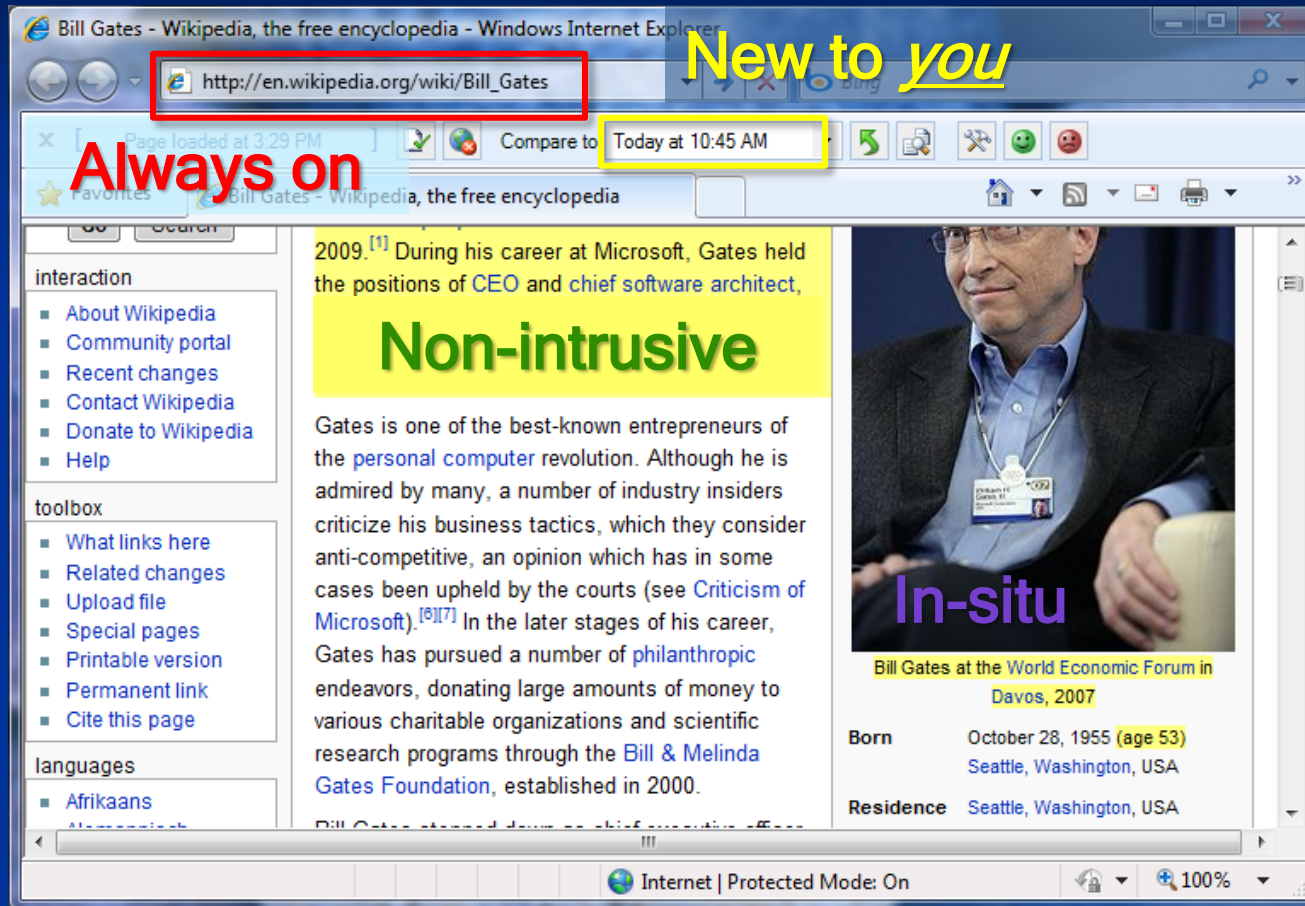
Diff-IE

Diff-IE toolbar



Changes to page since
your last visit

Interesting Features of Diff-IE



Try it: <http://research.microsoft.com/en-us/projects/diffie/default.aspx>

Examples of Diff-IE in Action

Expected New Content

HOME PAGE TODAY'S PAPER VIDEO MOST POPULAR TIMES TOPICS

Subscribe to The Times | Log In | Register Now

GUCCI **The New York Times** see what's inside

Saturday, February 5, 2011 Last Update: 1:27 PM ET

Search electric for orange

Follow Us f t

Subscribe to Home Delivery | Personalize Your Weather

Switch to Global Edition

JOBS
REAL ESTATE
AUTOS
ALL CLASSIFIEDS

WORLD
U.S.
POLITICS
NEW YORK
BUSINESS
DEALBOOK
TECHNOLOGY
SPORTS
SCIENCE
HEALTH
OPINION
ARTS
Books
Movies
Music
Television
Theater
STYLE
Dining & Wine
Fashion & Style
Home & Garden
Weddings/
Celebrations
TRAVEL
All Blogs
Cartoons
Classifieds
Corrections
Crossword /
Games
Education
First Look
Learning
Network
Multimedia

Obama Backs Suleiman-Led Transition

PHOTOGRAPHS OPINION VIDEO: KRISTOF AND FRIEDMAN



Hannibal Hanschke/European Pressphoto Agency

1 2 3 4 5 6 7 8 9 10 11 12 13 14

Protesters gathered in Tahrir Square, thousands of the demonstrators continued to occupy the square.

Elections Would be Held in September

By MARK LANDLER and KAREEM FAHIM 45 minutes ago

The Obama administration on Saturday backed attempts by Egypt's vice president, Omar Suleiman, to broker a compromise with opposition groups and prepare for elections.

NEWS ANALYSIS In Turkey's Example, Some See a Map for Egypt

By LONDON THOMAS Jr. 23 minutes ago

In politics and economics, Turkey provides a powerful illustration of success in the region.

WEEK IN REVIEW Seeing Secret Police's Methods Firsthand

By SOUAD MEKHENNET and NICHOLAS KULISH

For a day, we were trapped in a brutal maze that threw abuses of Egyptians into dreadful relief.

MORE FROM THE MIDDLE EAST


- **Egypt's Stability Hinges on Military** 3 minutes ago
- **Obama Said to Fault Mideast Intelligence**

OPINION »

OPINIONATOR | FIXES

Growing a Social Business

How the rural poor can start a business without taking on debt.



- Herbert: Bewitched by the Numbers | Comments (129)
- Collins: Siege of Planned Parenthood
- Blow: Kindling of Change
- Editorial: Court Politics
- Disunion: Women and Children First

BUSINESS DAY »

Cheesemakers Fret Over Possible New Rules

As the government reconsiders a 60-day aging rule for raw milk cheeses, artisans fear their businesses will suffer.



THE NEW DealB%k

dealbook.nytimes.com

Log In With Facebook

Log in to see what your friends are sharing on nytimes.com.
Privacy Policy | What's This?

WHAT'S POPULAR NOW

- The Siege of Planned Parenthood
- Obama Backs Suleiman-Led Transition

Book Review

Dining

Canvas for Pizza's New

Monitor

The screenshot shows a Twitter profile for Susan Dumais. The top navigation bar includes 'Home', 'Profile', 'Find People', 'Settings', 'Help', and 'Sign out'. The profile header shows 'Name Susan Dumais', '20 following', '12 followers', and '1 listed'. Below this, there are sections for 'Tweets' (0), 'Favorites', and 'Following' (a grid of 12 user avatars). The main content area is titled 'You follow 20 people' and lists several followed users with their profile pictures, names, and recent tweets. The users listed are pfromthenc, DARPA_News, huffingtonpost, nytimes, and dmRussell.

twitter Home Profile Find People Settings Help Sign out

You follow 20 people [add or invite more](#)

Name Susan Dumais
20 following 12 followers 1 listed

Tweets 0
Favorites

Following

User / Name	Actions
pfromthenc Loren Terveen Just had my first Surly Furious... waited until *after* I finished working on the proposal tonight. about 22 hours ago	
DARPA_News DARPA Arlington, VA Team standings from DARPA Network Challenge posted on http://bit.ly/5kdAZ1 about 4 hours ago	
huffingtonpost HuffingtonPost.com GOP Senator: We Will Unanimously Oppose Newest Health Care Compromise http://bit.ly/5xRBEI about 1 hour ago	
nytimes The New York Times New York, NY Citi Races to Pay Back Bailout Aid http://bit.ly/87E9Ry 5 minutes ago	
dmRussell CA, USA @roblyons Google Office is at 1101 New York Ave (although the entrance is on I) #gtadc about 13 hours ago	

Unexpected Important Content

Designated as an architectural masterpiece and home to a collection of one million books and other materials, you'll want to tour this Seattle treasure. Tours will begin at 8 p.m., an hour before the reception, to help ensure that everyone wanting a tour can be accommodated. Arrive at the 5th Avenue entrance of the library and wear your conference badge.

9:00 PM - 11:30 PM: Opening Reception / Seattle Public Library Norcliff Room

Opening Reception at the Norcliffe Room of the Central Library. Catered by Ray's Boathouse, celebrated Seattle restaurant and pioneer of Pacific Northwest Cuisine.

Wednesday, February 9

7:00 AM - 5:00 PM: Registration

8:45 AM - 10:30 AM: Native American Blessing / Welcoming led by Julian Argel (Washington) followed by [Keynote Address by Susan Dumais \(Microsoft Research\)](#)

Location: B Level, Courtyard Ballroom

Please note that the doors will close for a short time at the commencement of our ceremony out of deference to the Native American Blessing. Stragglers will be asked to wait until the blessing has concluded before being permitted to enter.

Session Chair: Jonathan Grudin (Microsoft Research)

10:30 AM - 4:00 PM: Special Event: Microsoft Visit


This half-day visit to the Microsoft campus in Redmond will involve presentations and demos by Microsoft Research staff, as well as tours at the Microsoft Corporate headquarters. Please note that it conflicts with iConference sessions 1 through 12.


UPDATE: this tour is officially full, and space is limited to attendees who already signed up at registration.


10:30 AM - 11:00 AM: Morning Break

11:00 AM - 12:30 PM:


Session 1: Alternative Event

 UNIVERSITY LIBRARIES
UNIVERSITY of WASHINGTON

 MORGAN & CLAYPOOL
PUBLISHERS

 texifter

iConference 2011 is hosted by
the University of Washington
Information School

 Information School
UNIVERSITY of WASHINGTON

Understand Page Dynamics

The image shows a screenshot of a Bing search results page for the query "jaime teevan". The page layout includes the Bing logo, search filters, related searches, and a list of search results. Several elements are highlighted in yellow to illustrate page dynamics.

bing
jaime teevan

Web Images

RELATED SEARCHES
Susan Dumais

ALL RESULTS 1-10 of 11,700 results · Advanced

Sponsored sites

We Found Jaime Teevan
Instant-Address, Phone, Age & More.
Search for Jaime Teevan Now!
www.Intelius.com
See your message here

Jaime Teevan, Ph.D.
Jaime Teevan, Ph.D. Researcher studying information retrieval and human computer interaction at Microsoft Research.
research.microsoft.com/en-us/um/people/teevan · Cached page

Jaime Teevan: Work
Jaime Teevan: Doctoral candiate at Massachusetts Institute of Technology. Research in information retrieval and information architecture.
people.csail.mit.edu/teevan/work · Cached page

DBLP: Jaime Teevan
2010; 36 : Jaime Teevan, Susan T. Dumais, Daniel J. Liebling: A longitudinal study of how highlighting web content change affects people's web interactions.
www.informatik.uni-trier.de/~ley/db/indices/a-tree/t/Teevan:Jaime.html · Cached page

Jaime Teevan - Pipl Profile
Pipl profile of Jaime Teevan. Quick facts, personal profiles, publications, contact details and much more.
pipl.com/directory/people/Jaime/Teevan · Cached page

Jaime Teevan: Work
Jaime Teevan: Doctoral candiate at Massachusetts Institute of Technology. Publications.
people.csail.mit.edu/teevan/work/publications/subject.html · Cached page

Jaime Teevan - LinkedIn
Research · 232 connections · Greater Seattle Area
View Jaime Teevan's professional profile on LinkedIn. LinkedIn is the world's largest business network, helping professionals like Jaime Teevan discover inside connections to ...
www.linkedin.com/pub/jaime-teevan/0/542/7ab · Cached page

TR35: Jaime Teevan, 32 - Technology Review
From MIT. Information on Emerging Technologies & impact on business & society

Expected



Expected New Content



Monitor



Unexpected Important Content

33 - Estimate Mothers & More Expo... Use Chapter Names and Events for more information. Yeah, baby!

Subject	Replies	Views	Last post
Small Business + ...	21	206	Monday 16, 2008, 9:00 AM
Coaching Inquiries	10	83	Monday 16, 2008, 9:00 AM
Back to work Part II	8	43	Monday 16, 2008, 9:00 AM
Aldergate Christian Preschool - HIKING	0	17	Monday 16, 2008, 9:00 AM
do you know anyone in Italy or Sweden?	3	20	Monday 16, 2008, 9:00 AM
Call an Interview and I'm Recalling ...	94	562	Monday 16, 2008, 9:00 AM
Autism and Industry? Finding the Right Fit	7	105	Monday 16, 2008, 9:00 AM
Any members work at Amazon.com?	10	118	Monday 16, 2008, 9:00 AM
Time Magazine on SAHM - getting back into workforce	1	75	Monday 16, 2008, 9:00 AM
Freemasonry Resurgence	12	41	Monday 16, 2008, 9:00 AM
Congrat designer or Admin?	0	45	Monday 16, 2008, 9:00 AM

Attend to Activity



Serendipitous Encounter

Unexpected



Edit



Understand Page Dynamics



Unexpected Unimportant Content

Studying Diff-IE

■ Feedback buttons



■ Survey

- Prior to installation
- After a month of use

6. How often do you find the following types of pages change?

Required	Always	Often	Sometimes	Rarely	Never
News pages	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Message boards, forums, newsgroups	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Company homepages	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal homepages of people you know	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pages with product information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reference pages (dictionaries, yellow pages, maps)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wikipedia pages	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Blips you read	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Search engine results	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

■ Logging

- URLs visited
- Amount of change when revisited



■ Experience interview






In situ
Representative




Experience

Longitudinal

People Revisit More

- Perception of revisitation remains constant
 - How often do you revisit? 
 - How often are revisits to view new content? 
- Actual revisitation increases
 - Last week: 45.0% of visits are revisits 
 - First week: 39.4% of visits are revisits
- Why are people revisiting more with DIFF-IE?

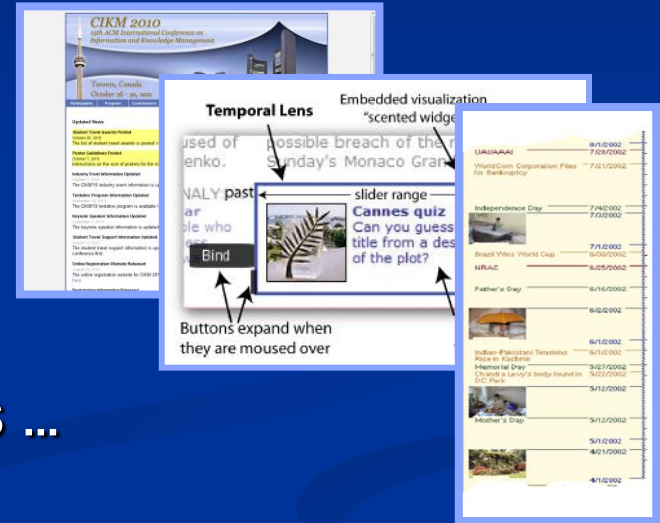
Revisited Pages Change More

- Perception of change increases
 - What proportion of pages change regularly?  8%
 - How often do you notice unexpected change?  17%
- Amount of change seen increases
 - Last week: 32.4% revisits changed, by 9.5%  51+%
 - First week: 21.5% revisits changed, by 6.2%
- Diff-IE is driving visits to changed pages
 - It supports people in understanding change

Other Examples of Dynamics and User Experience

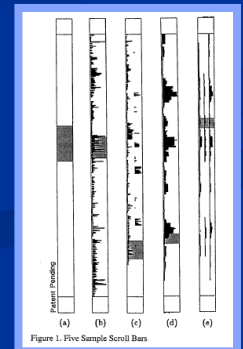
■ Content changes

- Diff-IE (Teevan et al., 2008)
- Zoetrope (Adar et al., 2008)
- Diffamation (Chevalier et al., 2010)
- Temporal summaries and snippets ...



■ Interaction changes

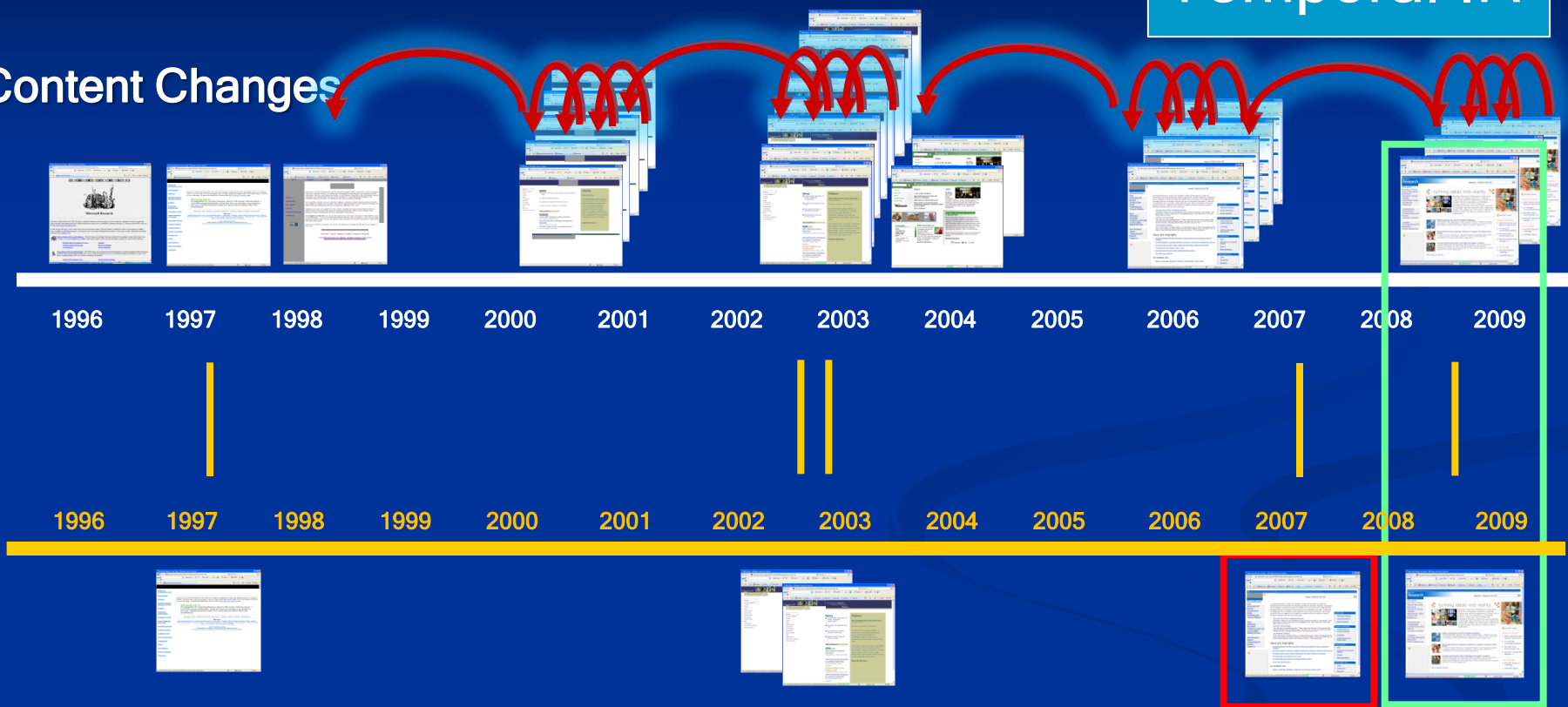
- Explicit annotations, ratings, wikis, etc.
- Implicit interest via interaction patterns
 - Edit wear and read wear (Hill et al., 1992)



Leveraging Dynamics for Retrieval

Temporal IR

Content Changes



User Visitation/ReVisitation

Temporal Retrieval Models

- Current retrieval algorithms look only at a single snapshot of a page
- But, Web pages change over time
- Can we can leverage this to improved retrieval?
 - Pages have different *rates of change*
 - Different priors (using change vs. link structure)
 - Terms have *different longevity (staying power)*
 - Some are always on the page; some transient
 - Language modeling approach to ranking

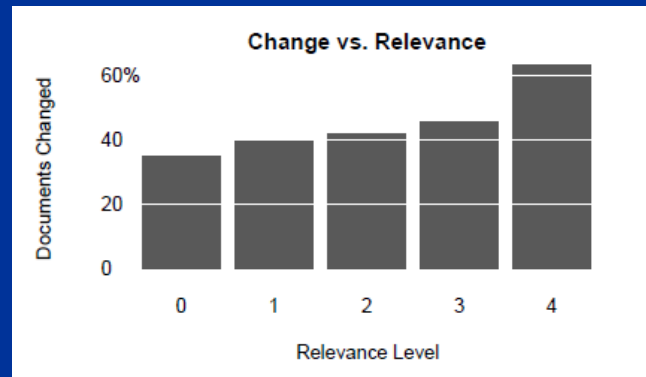
$$P(D|Q) = P(D) \cdot P(Q|D)$$

Change prior

Term longevity

Relevance and Page Change

- Page change is related to relevance judgments
 - Human relevance judgments
 - 5 point scale - Perfect/Excellent/Good/Fair/Bad
 - Rate of Change -- 60% Perfect pages; 30% Bad pages



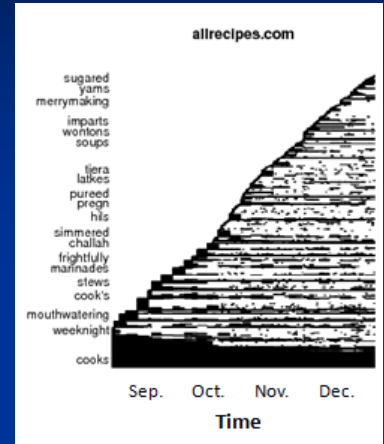
- Use change rate as a document prior (vs. priors based on link structure like Page Rank)
 - Shingle prints to measure change

$$P(D|Q) = P(D) \cdot P(Q|D)$$

Change prior

Relevance and Term Change

- Terms patterns vary over time
- Represent a document as a mixture of terms with different “staying power”
 - Long, Medium, Short



$$P(Q | D) = \lambda_L P(Q | D_L) + \lambda_M P(Q | D_M) + \lambda_S P(Q | D_S)$$

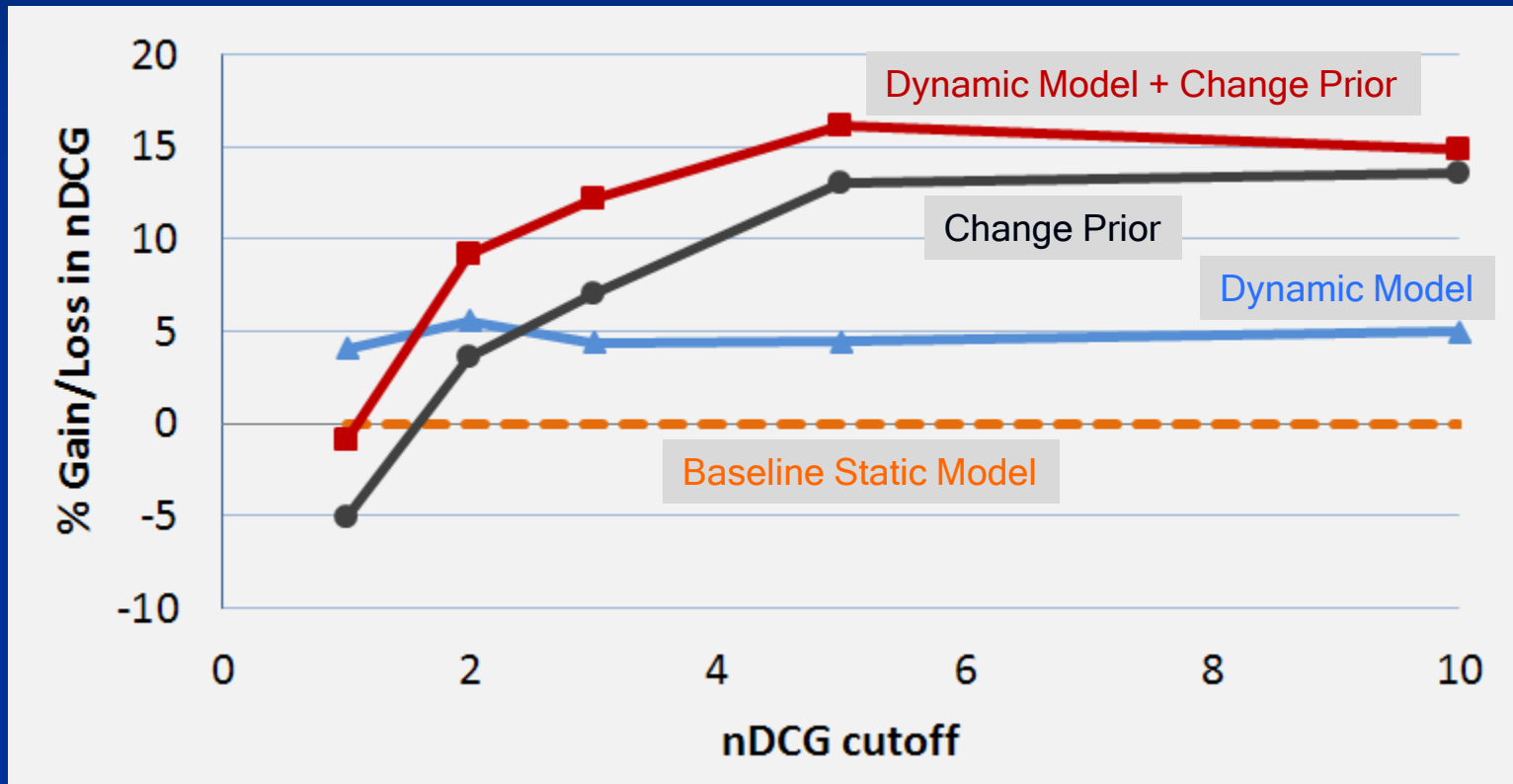
$$P(D | Q) = P(D) \cdot P(Q | D)$$

↑
Term longevity

Evaluation: Queries & Documents

- 18K Queries, 2.5M Judged Documents
 - 5-level relevance judgment (Perfect ... Bad)
- 2.5M Documents crawled weekly for 10 wks
- Navigational queries
 - 2k queries identified with a “Perfect” judgment
 - Assume these relevance judgments are consistent over time

Experimental Results

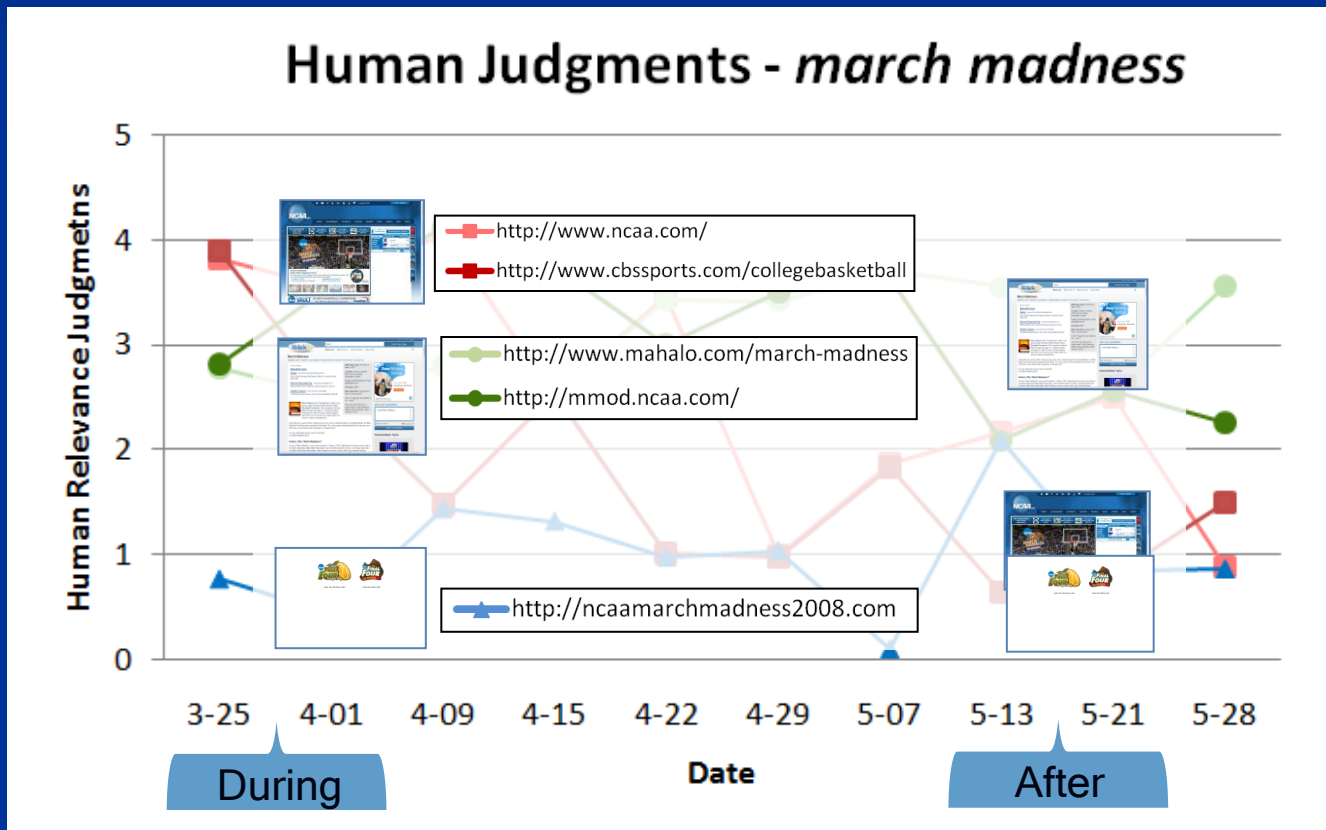


Temporal Retrieval, Ongoing Work

- Initial evaluation
 - Focused on navigational queries
 - Assumed their relevance is “static” over time
- But, there are many other cases ...
 - E.g., *US Open 2010* (in June vs. Sept)
 - E.g., *World Cup Results* (in 2010 vs. 2006)
- Ongoing evaluation
 - Collecting explicit relevance judgments, query frequency, interaction data and page content over time
 - Developing temporal IR models, temporal snippets

Relevance over Time

- Query: *march madness* [Mar 15 - Apr 4, 2010]



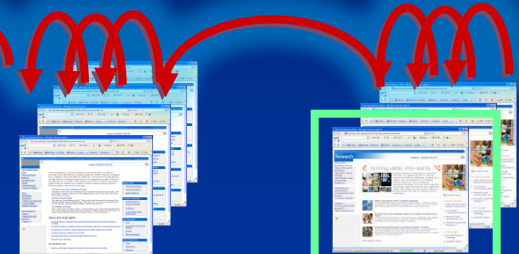
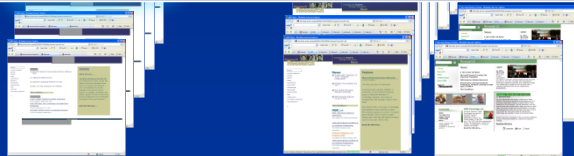
Other Examples of Dynamics and Information Systems

- Query dynamics
 - Kulkarni et al. (2011); Jones & Diaz (2004); Diaz (2009); Kotov et al. (2010)
- Document dynamics, for crawling and indexing
 - Adar et al. (2009); Cho & Garcia-Molina (2000); Fetterly et al. (2003)
- Temporal retrieval models
 - Elsas & Dumais (2010); Liu & Croft (2004); Efron (2010); Aji et al. (2010)
- Extraction of temporal entities within documents
- Protocol extension for retrieving versions over time
 - E.g., Memento (Van de Sompel et al., 2010)

Summary

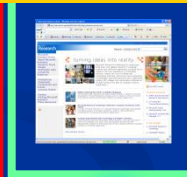
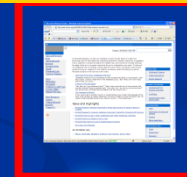
Temporal IR:
Leverages change
for improved IR

Web content changes: page-level, term-level



Relating revisitation and change allows us to

- Identify pages for which change is important
- Identify interesting components within a page



People revisit and re-find Web content

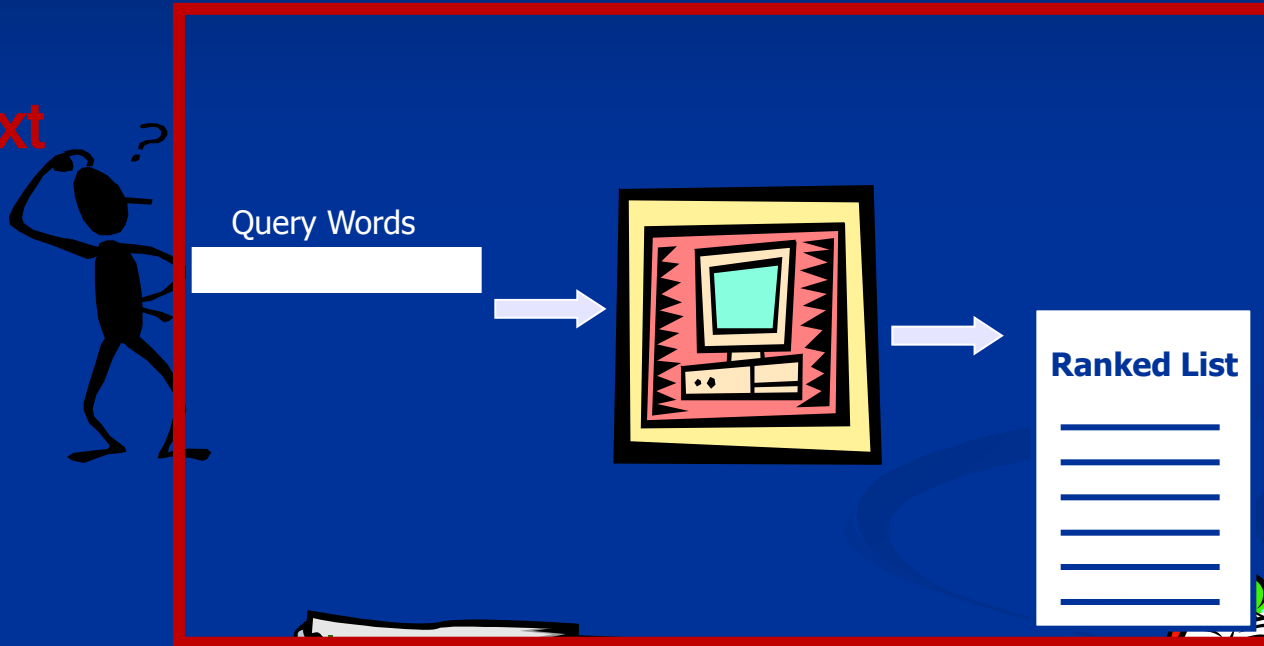
Diff-IE: Supports (and influences) interaction and understanding

Challenges and Opportunities

- Temporal dynamics are pervasive in information systems
- Influence many aspect of information systems
 - Systems: protocols, crawling, indexing, caching
 - Document representations: meta-data generation, information extraction, sufficient statistics at page and term-level
 - Retrieval models: term weights, document priors, etc.
 - User experience and evaluation
- Better supporting temporal dynamics of information
 - Requires digital preservation and temporal metadata extraction
 - Enables richer understanding of the evolution (and prediction) of key ideas, relations, and trends over time
- Time is one important example of context and IR
 - Others include: location, individuals, tasks ...

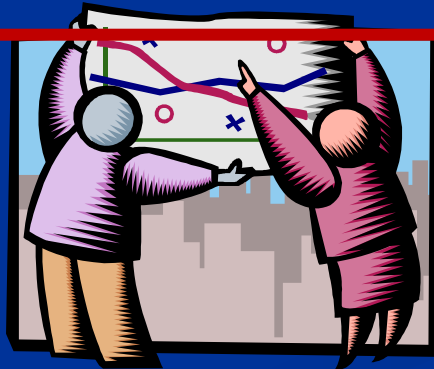
Think Of Search (Search) Boxes

User
Context



Document
Context

Task/Use
Context



Feb 2011

Feb 2005

Feb 2000

Harry Bruce
Home Publications Research Teaching Services Curriculum Vitae

Harry Bruce is the Dean of the University of Washington Information School. His research and teaching focus on human-computer interaction, information seeking and use, and personal information management in individual information environments. Dr. Bruce's research has been funded by the National Science Foundation, the Institute of Management and Labor Services (IMLS), the National Endowment for the Humanities, the Department of Engineering Education and Training.

Recent Presentations

- Bruce H. "The Academic's View of Information Needs." 2010 International Symposium on the Information Science and Technology of Library and Information Science, November 10-17, 2010 Taipei, Taiwan.
- Bruce H. "The Emerging Role of Librarians in an Information Age." Friends of the Seattle Public Library, University of Washington, Seattle, WA, March 12, 2010.
- Bruce H. "Academics and Higher Education: Address and the Use of Online Catalogs." IFLA Academic Library Director Symposium, Portland OR, November 16, 2009.
- **Emerging Issues: Things Found and Lost.** University of Colorado, University of Michigan, University of Wisconsin, University of Washington, Portland OR, November 16, 2009.

Dean and Professor
The Information School
University of Washington

Information School
News & Events Programs Courses People Resources Research Technology Links

Mike Eisenberg
Dean Emeritus & Professor
University of Washington
Information School

Mike Eisenberg is a leading expert in information science, information technology, and information management in learning and teaching. He focuses on the use of information and information technology by individuals and organizations to meet their information needs and manage their information more effectively and efficiently. He is nationally known for his innovative approach to information problem-solving and technology in learning and teaching. Dr. Mike is in great demand as a speaker and delivering numerous keynote addresses and workshops.

He earned his BA and MS from SUNY at Albany and his Ph.D. in Information Transfer from the School of Information Studies at Syracuse University.

Personal Motto: "Make it better."

Dr. Michael B. Eisenberg conducts research, writes, consults, and lectures frequently on:

- Information library
- Information technology
- Information management in learning and teaching
- Information and library education.

Harry Bruce

Assistant Dean, Research
Associate Professor
The Information School
University of Washington

Recent Presentations

- **Trustworthy Information: A Review of the IASL Conference**
- **Trustworthy Information: A Review of the IASL Conference**

Recent Publications

- **Trustworthy Information: A Review of the IASL Conference**
- **Trustworthy Information: A Review of the IASL Conference**

Recent Presentations

- **Trustworthy Information: A Review of the IASL Conference**
- **Trustworthy Information: A Review of the IASL Conference**

CV

- **Trustworthy Information: A Review of the IASL Conference**
- **Trustworthy Information: A Review of the IASL Conference**

School of Library and Information Science
News & Events Programs Courses People Resources Research Technology Links

Dr. Michael B. Eisenberg
Director
School of Library and Information Science
Box 353200
Seattle WA 98195-3200
206.543.5744
mbe@u.washington.edu

Dr. Michael B. Eisenberg received his Ph.D. from SUNY at Albany and his Ph.D. in Information Transfer from the School of Information Studies at Syracuse University.

He earned his M.S. from SUNY at Albany and his Ph.D. in Information Transfer from the School of Information Studies at Syracuse University.

Harry Bruce, Associate Professor
E-mail: hbruce@u.washington.edu
Office: 3700 Mary Gates Hall
Phone: (206) 534-0965

Personal Web Page:
<http://web.archive.org/web/2005031611127/http://www.ischool.uwashington.edu/harry/>

Profile:

Dr. Harry Bruce's teaching, research and publishing are centered on human factors in communications and information technology. In recent years, his research has focused primarily on networked information environments and the Internet. Associate Professor Bruce has an M.A.B. (by research, 1993) and a Ph.D. (1995) from the University of New South Wales (UNSW), Australia, and is currently an outstand...

Dr. Michael B. Eisenberg
Dean
University of Washington
Information School
Suite 320 Mary Gates Hall
Seattle WA 98195-3200
206.543.5744
mbe@u.washington.edu

Mike conducts research, writes, consults, and lectures frequently on:

Information problem-solving:

- Information library
- Information technology
- The Internet
- Information management in learning and teaching.

He focuses on the use of information and information technology by individuals and organizations to meet their information needs and manage their information more effectively and efficiently. He is nationally known for his innovative approach to problem-solving and critical thinking/information skills development. Mike is in great demand as a speaker and holds numerous workshops across the country on the high skills approach to information problem-solving and technology in learning and teaching.

Mike earned his M.S. from SUNY at Albany and his Ph.D. in Information Transfer from the School of Information Studies at Syracuse University.

Some Interesting Articles by about Mike:

- **"It's All About Learning: Ensuring That Students Are Effective Users of**

iConference 2011
Seattle, WA

iConference 2011 Home
Registration is Now Open

Use This Site to Register. Please Note that the Final Registration Deadline is **January 24, 2011**. See our [Registration Site for details](#).

What's New on this site:

- **What's New on this site:** Please Note that the Final Registration Deadline is January 24, 2011. See our Registration Site for details.
- **What's New on this site:** Please Note that the Final Registration Deadline is January 24, 2011. See our Registration Site for details.

Diff-IE ... try it!
<http://research.microsoft.com/en-us/projects/diffie/default.aspx>