

Temporal Dynamics and Information Retrieval

Susan T. Dumais

Microsoft Research

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

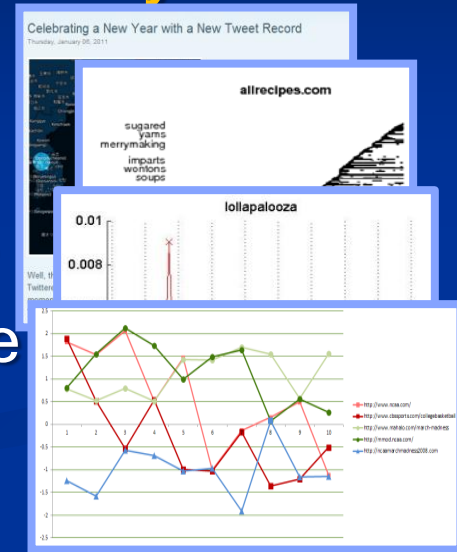
In collaboration with:

Eric Horvitz, Jaime Teevan, Eytan Adar, Jon Elsas, Dan Liebling,
Richard Hughes, Krysta Svore, Kira Radinsky

Change is Everywhere in IR

- Change is everywhere in digital information systems

- New documents appear all the time
- Document content changes over time
- Queries and query volume change over time
- What's relevant to a query changes over time
 - E.g., *U.S. Open 2012* (in June vs. Sept)
- User interaction changes over time
 - E.g., anchor text, “likes”, query-click streams, social networks, etc.
- Relations between entities change over time
 - E.g., President of the US is \leftrightarrow [in 2008 vs. 2004 vs. 2000]



- Change is pervasive in digital information systems
... yet, most retrieval systems ignore it !

Digital Dynamics Easy to Capture

- Easy to capture
- But ... few tools or algorithms support dynamics

Susan Dumais



LBH 9/25/96

SeniorPrincipal Researcher, [Decision Theory & Adaptive Systems & Interaction Group](#),
[Microsoft Research](#)

E-mail: sdumais@microsoft.com

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

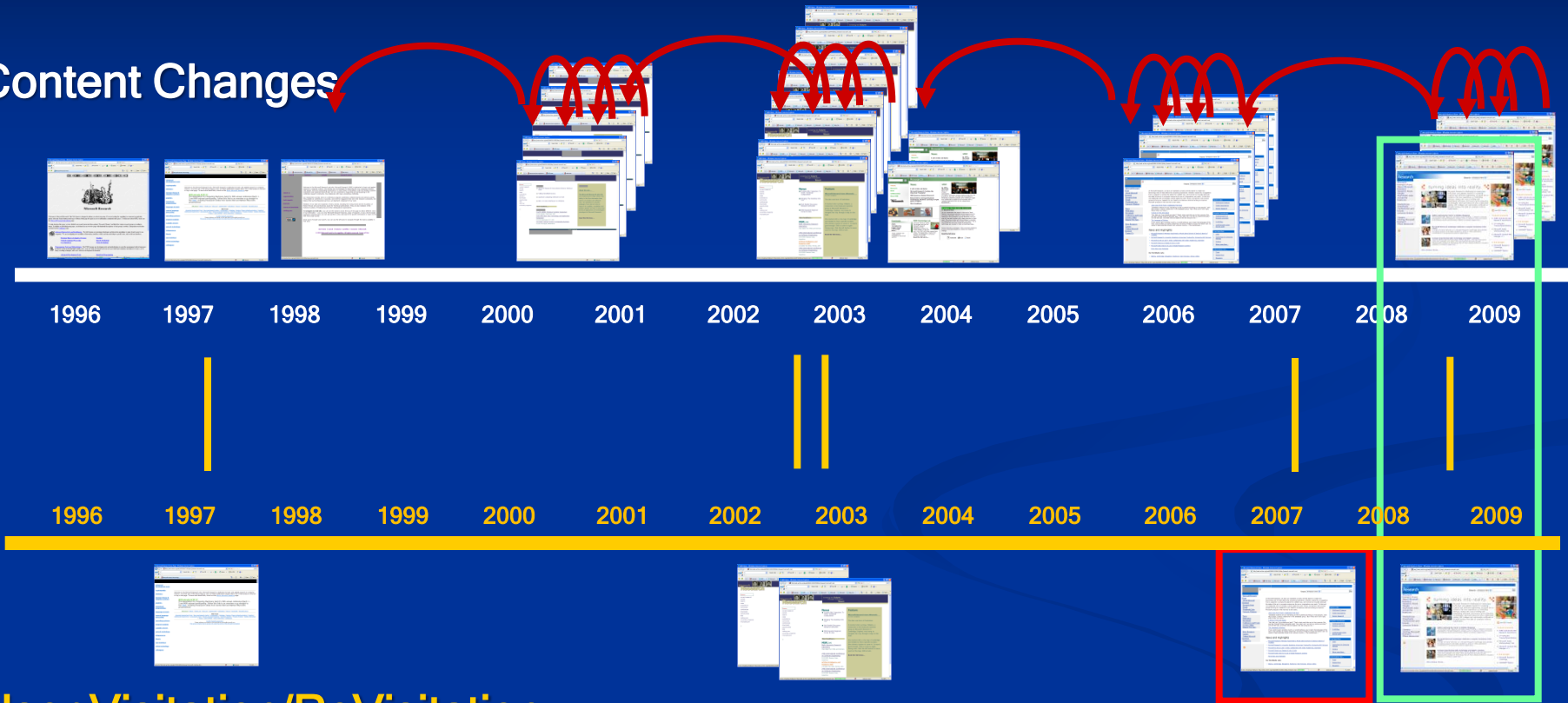
Research Activities:

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: ~~text~~personal information management, web search, question answering, information retrieval ~~and, text~~ categorization, collaborative filtering, interfaces for ~~combining~~improved 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 ~~Bellcore LSI page~~, ~~Bellcore~~ (now ~~Telcordia~~) LSI page.

Web Dynamics

Content Changes



User Visitation/ReVisitation

Today's Browse and Search Experiences

But, ignores ...

Overview

■ Change on the Web

- Content changes over time
- User interaction varies over time (queries, re-visitation, anchor text, query-click stream, “likes”)
- Tools for understanding Web change (e.g., Diff-IE)

■ Improving Web retrieval using dynamics

- Query trends over time
- Retrieval models that leverage dynamics
- Task evolution over time

Overview

■ Change on the Web

- Content changes over time
- User interaction varies over time (queries, re-visitation, anchor text, query-click stream, “likes”)
- Tools for understanding Web change (e.g., Diff-IE)

■ Improving Web retrieval using dynamics

- Query trends over time
- Retrieval models that leverage dynamics
- Task evolution over time

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.

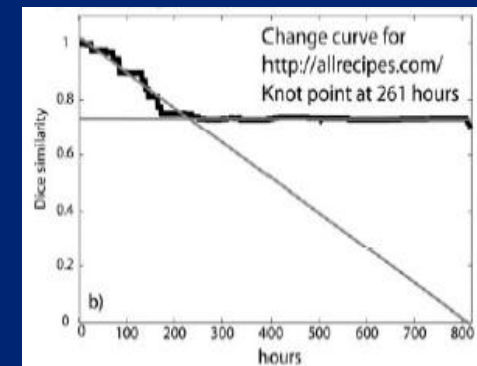
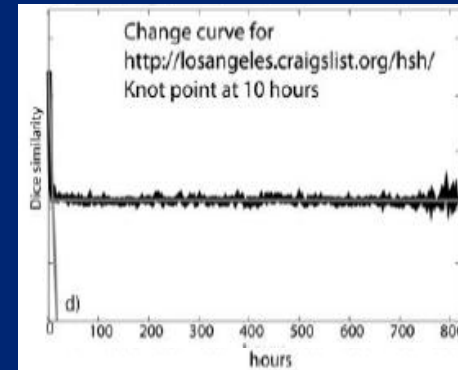
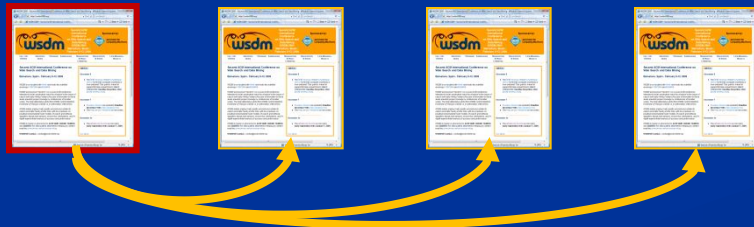
■ Amount of change

■ Time between changes

- Avg. Dice coeff. = 0.80
- Avg. time bet. change = 123 hrs.
- .edu and .gov pages change infrequently, and not by much
- .com pages change at an intermediate rate, but by a lot
- 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



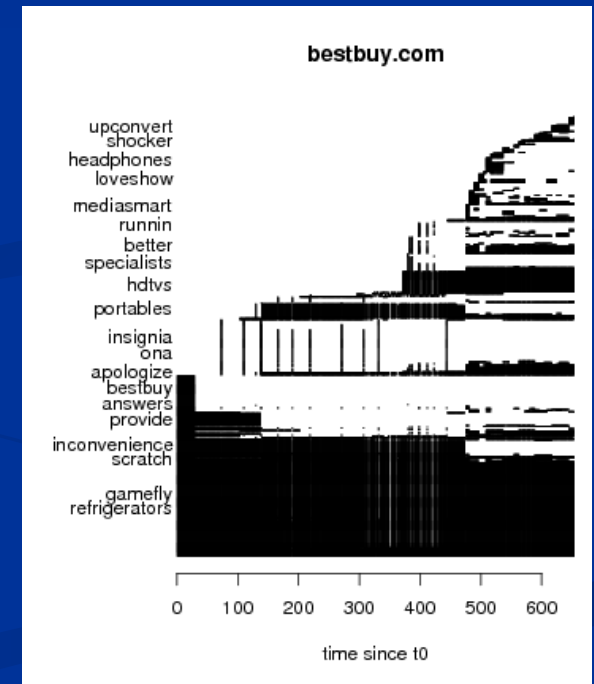
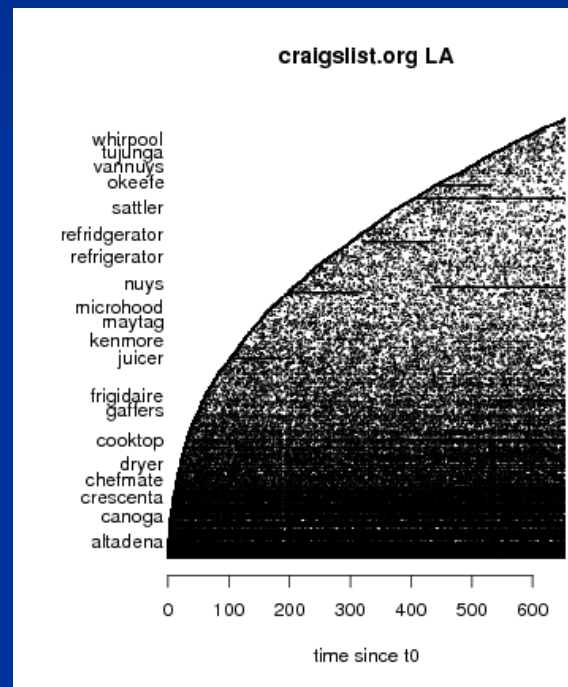
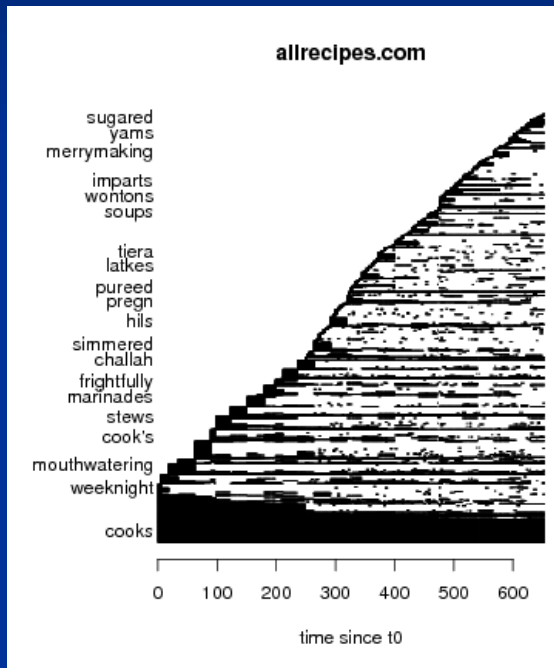
Measuring Within-Page Change

- 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



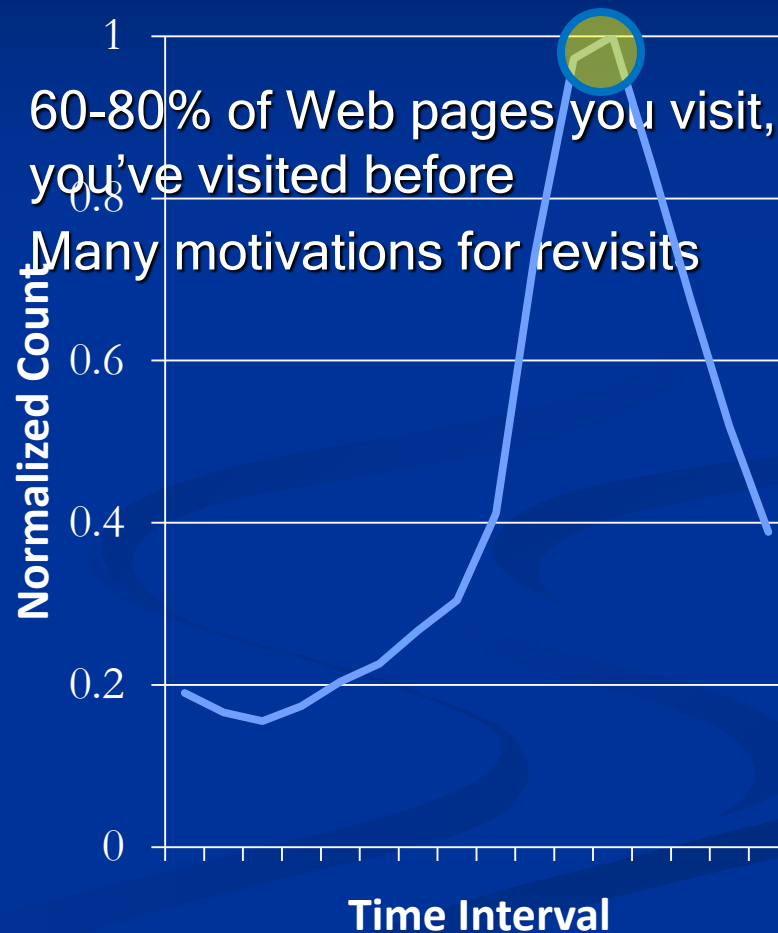
User Visitation/ReVisitation

What was the last Web page you visited?

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

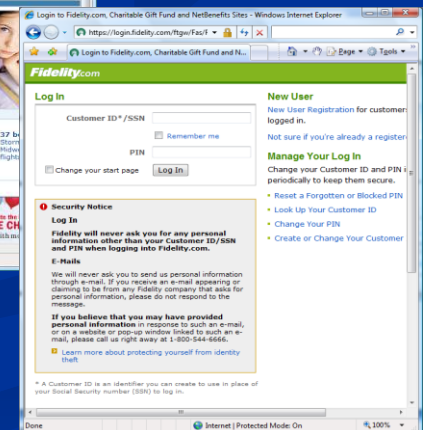
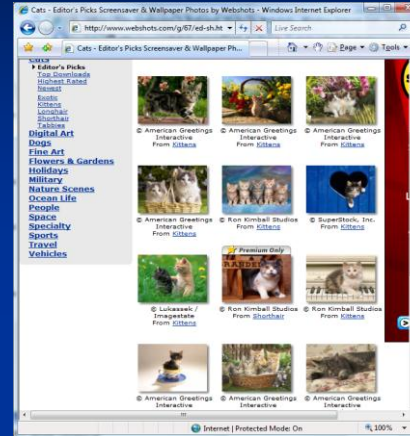
Measuring Revisitation

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



Four Revisitation Patterns

- *Fast*
 - Hub-and-spoke
 - Navigation within site
- *Hybrid*
 - High quality *fast* pages
- *Medium*
 - Popular homepages
 - Mail and Web applications
- *Slow*
 - Entry pages, bank pages
 - Accessed via search engine



Relationships Between Change and Revisitation



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

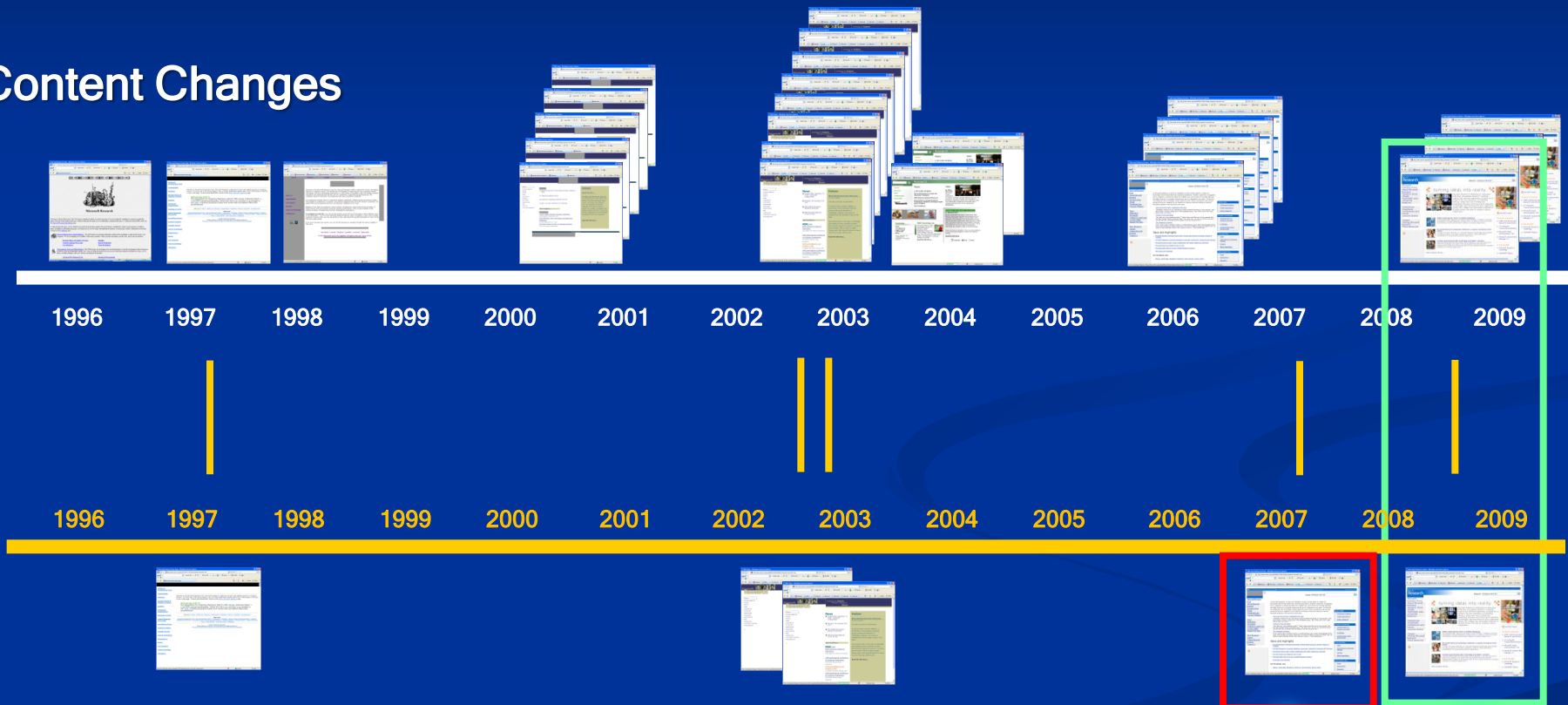
Revisitation and Search (Re-finding)

- 60-80% of the Web page visits are re-revisits
- 33-43% of queries are re-finding
 - Repeat query (33%)
 - Q: *microsoft research*
 - Click same or different URLs
 - Repeat click (39%)
 - <http://research.microsoft.com/>
 - Q: *microsoft research; msr*
 - Big opportunity (43%)

		Repeat Click	New Click
Repeat Query	33%	29%	4%
New Query	67%		

Building Support for Web Dynamics

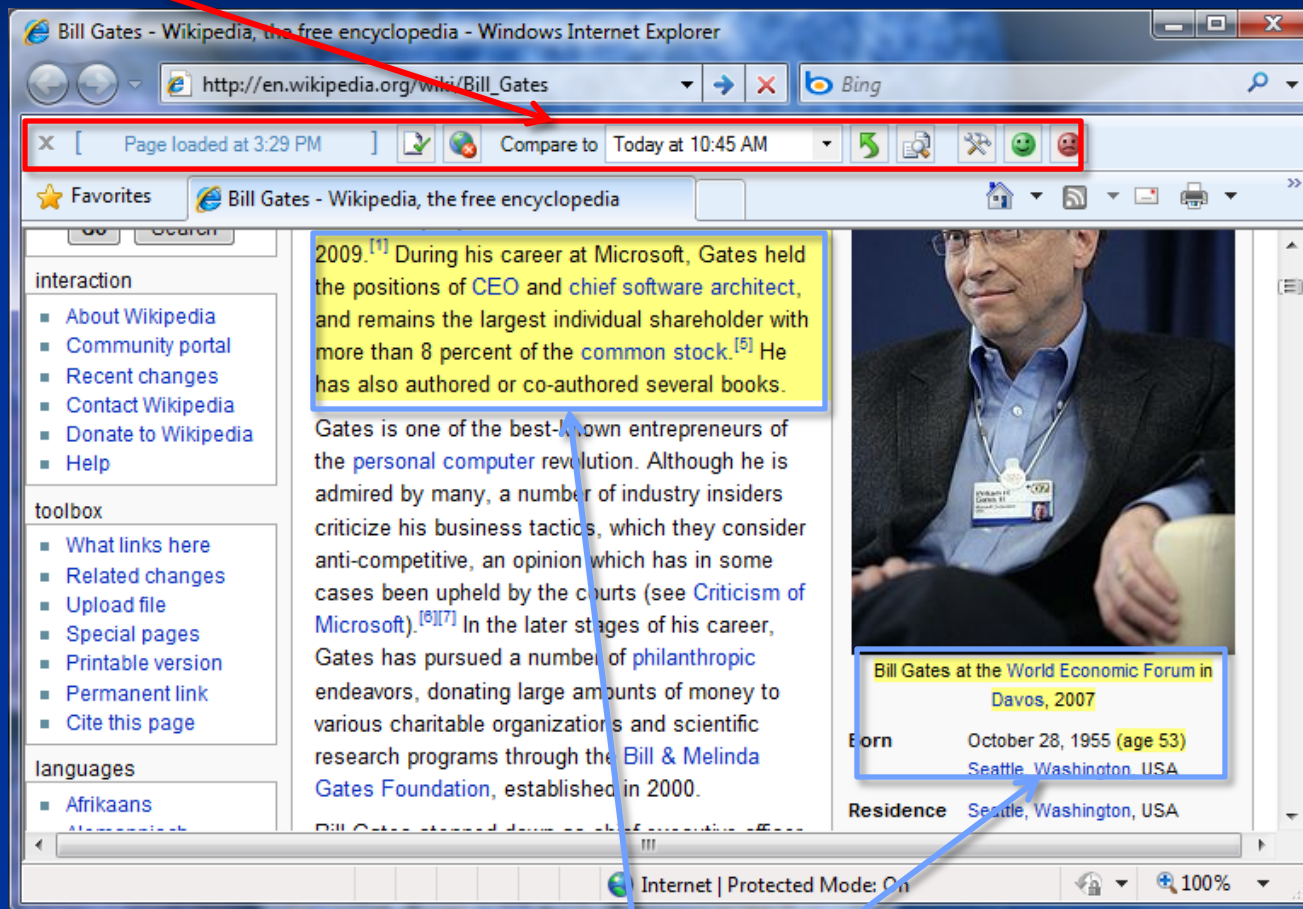
Content Changes



User Visitation/ReVisitation

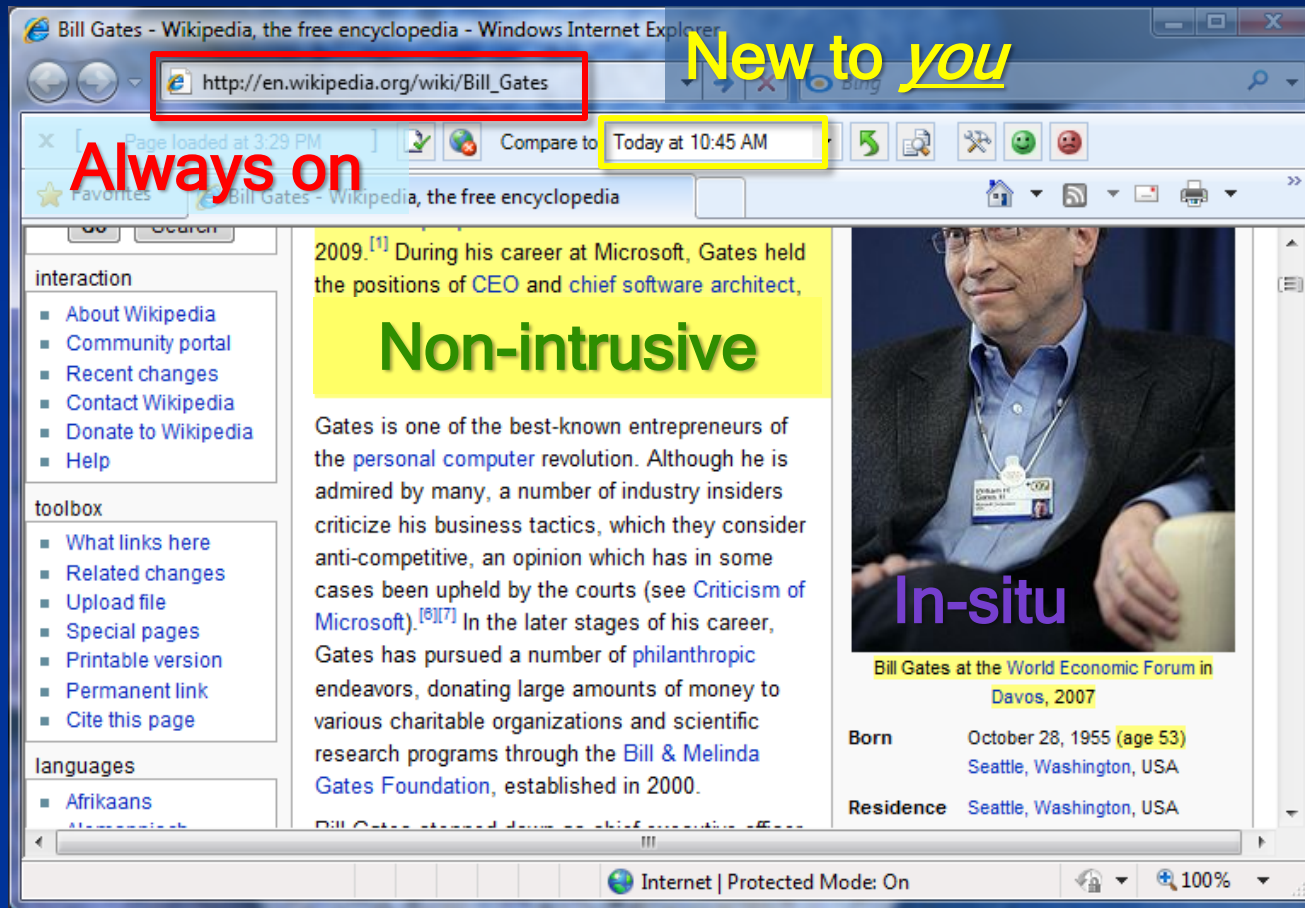
Diff-IE

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](#) Edition: U.S. / Global [Subscribe: Digital / Home Delivery](#) [Log In](#) [Register Now](#)

SHOP NOW AT
MARCJACOBS.COM

The New York Times

Tuesday, April 24, 2012 Last Update: 5:21 PM ET

Get a Full Times Experience.

[BECOME A DIGITAL SUBSCRIBER](#)

[ING DIRECT](#)

Follow Us [Subscribe to Home Delivery](#) | [Personalize Your Weather](#)

WORLD
U.S.
POLITICS
NEW YORK
BUSINESS
DEALBOOK
TECHNOLOGY
SPORTS
SCIENCE
HEALTH
ARTS
STYLE
OPINION

[Autos](#)
[Blogs](#)
[Books](#)
[Cartoons](#)
[Classifieds](#)
[Crosswords](#)
[Dining & Wine](#)
[Education](#)
[Event Guide](#)
[Fashion & Style](#)
[Home & Garden](#)
[Jobs](#)
[Magazine](#)
[Movies](#)
[Music](#)
[Obituaries](#)
[Real Estate](#)
[Spectrum](#)

Debt Collector Is Faulted for Tough Tactics in Hospitals

By JESSICA SILVER-GREENBERG 4:03 PM ET

One of the nation's largest medical debt-collection companies is under fire in Minnesota for placing employees in hospitals and demanding patients pay before receiving treatment, documents show.

[Post a Comment](#) | [Read \(308\)](#)

Hints of Collusion Between News Corp. and British Minister

By JOHN F. BURNS and ALAN COWELL 2:45 PM ET

Evidence presented at an inquiry suggested that the culture minister, or an aide claiming to speak for him, worked covertly to help win approval for a takeover of the BSkyB network.

[Post a Comment](#) | [Read \(11\)](#)

Drew Litton for The New York Times

Late Innings for the Sports Cartoon

By RICHARD SANDOMIR 4:12 PM ET

Even before the newspaper industry started shrinking, editors began to view sports cartoonists as vestiges of a bygone era. Above, Drew Litton's take on the issue.

CAMPAIGN 2012

THE CAUCUS

With Primary Field, Romney Looks to Win Big

By MICHAEL D. SHEAR

Five states vote Tuesday in primaries that party leaders hope will cement Mitt Romney's status as the nominee.

- [After Delaware Primary, Gingrich Will 'Reassess' Race](#) 5:18 PM ET
- [Obama Calls Agents in Scandal 'Knuckleheads'](#) 4:20 PM ET

OPINION

Borderlines: One Island, Two Countries
Divided islands, like Market in the Baltic Sea, conform to a version of Sayre's law: the smaller the territory, the more confusing the border.

- [Brooks: Creative Monopoly](#)
- [Bruni: Usual Scapegoats](#)
- [Editorial: France Votes](#)
- [Campaign Stops: The Mega Millions Solution](#)
- [Draft: Is Texting Writing?](#)
- [Loyal Opposition: Romney and the N.Y. Primary](#)

MARKETS » At close 04/24/2012

S.&P. 500	Dow	Nasdaq
1,371.97	13,001.56	2,961.60
+5.03	+74.39	-8.85
+0.37%	+0.58%	-0.30%

[GET QUOTES](#) [My Portfolios](#)
Stock, ETFs, Funds

TRY IT NOW
4 WEEKS FOR 99¢
[CLICK HERE](#)

BREAKING NEWS & OPINION
24/7 UPDATES THE CAUCUS BLOG

POLITICAL
THE ON

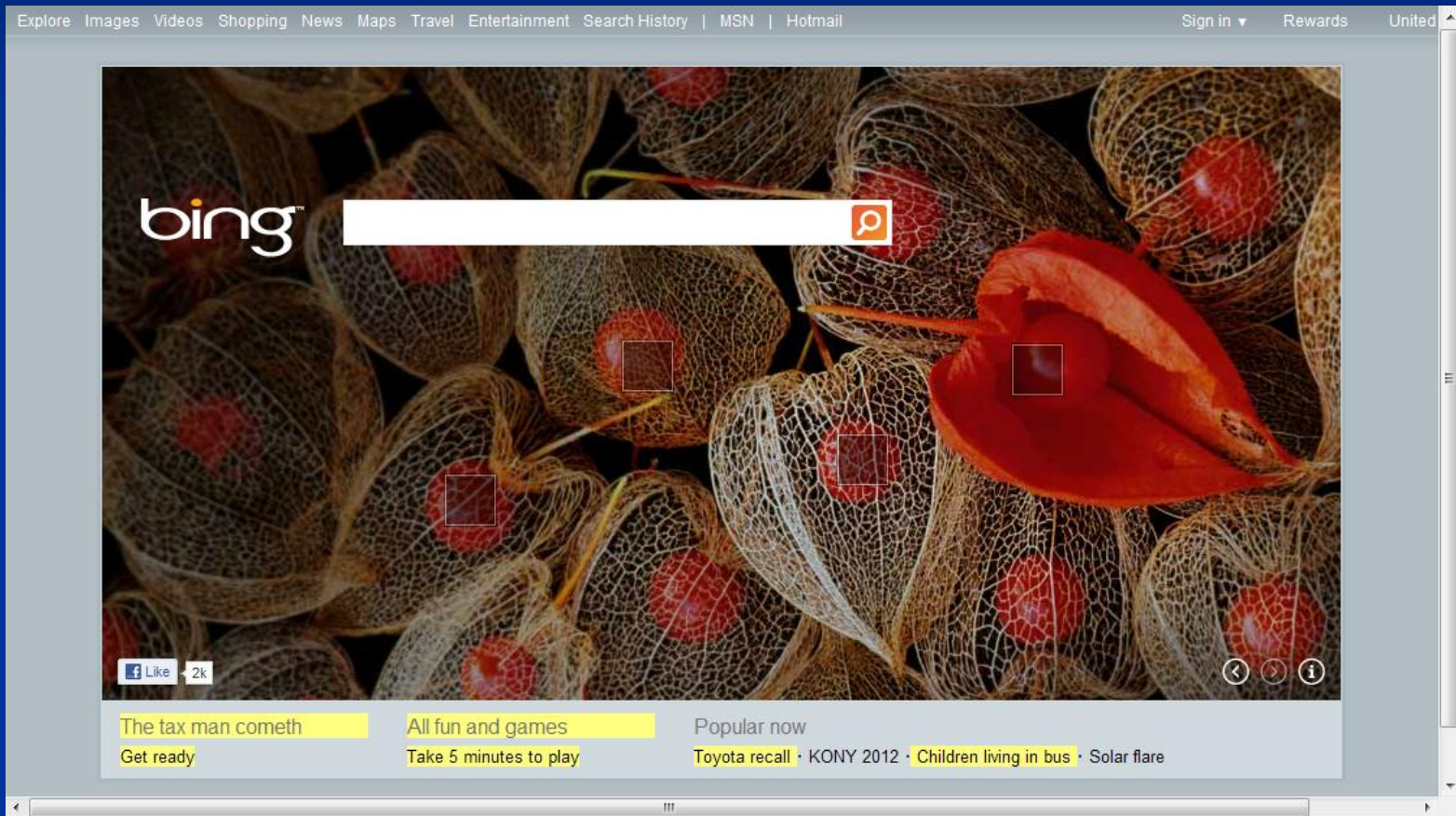
SIAM-SDM 2012, Apr 27

Monitor

The screenshot shows a Twitter profile for Susan Dumais. The header includes the Twitter logo and navigation links: Home, Profile, Find People, Settings, Help, and Sign out. The profile section displays the name 'Susan Dumais' and statistics: 20 following, 12 followers, and 1 listed. Below this are tabs for Tweets (0), Favorites, and Following. The Following tab is active, showing a grid of 12 user avatars. The main content area, titled 'You follow 20 people', lists the following users:

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	

Serendipitous Encounters



Unexpected Important Content



Understand Page Dynamics

The screenshot shows a Bing search results page for the query "jaime tee van". The page layout includes a search bar at the top with the query and a "Web" tab selected. Below the search bar, there are tabs for "Web" and "Images". The search results are displayed in a list format, with each result including a title, a brief description, and a URL. The results are sorted by relevance, and the first result is "Jaime Teevan, Ph.D.". The page also features a "RELATED SEARCHES" section on the left, showing "Susan Dumais". On the right, there is a "Sponsored sites" section with a link to "We Found Jaime Teevan". The page is displayed in a browser window with a standard Windows taskbar at the bottom.

bing™

jaime tee van

Web Images

RELATED SEARCHES

Susan Dumais

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

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 candidate 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 candidate 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

Sponsored sites

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

Expected



Expected New Content



Monitor

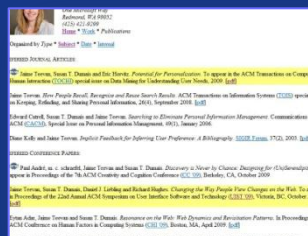


Unexpected Important Content

A screenshot of a forum thread. The title is "Katharine's More Info". The thread contains several posts with details about Katharine's work and activities. The posts are organized by date and time.

Subject	Started by	Replies	Views	Last post
Small Business - 1111	Della Davis	21	204	Monday, 11/11/11
Outgoing Inquiries	MRS Morris	0	83	Monday, 11/11/11
Back to work Part Time	Bob Morrison	0	43	Monday, 11/11/11
Adoptive Christian Preschool - KIDING	Heather Rose	0	17	Monday, 11/11/11
Do you know anyone in Italy or Sweden?	Virginia Manning	0	30	Monday, 11/11/11
Get an interview and the feeling well...	Theresa Dwyer	0	82	Monday, 11/11/11
Available in Indiana? Finding this right	James Taylor	7	105	Monday, 11/11/11
Any members work at Amazon.com?	Katharine Davis	0	108	Monday, 11/11/11
Time Magazine on SAT/ACT getting back into	Katharine Davis	0	70	Monday, 11/11/11
Frederick Rosemont	Katharine Davis	0	45	Monday, 11/11/11
Google's designer or Admin?	Heather Rose	0	45	Monday, 11/11/11

Attend to Activity



Serendipitous Encounter

Unexpected



Edit



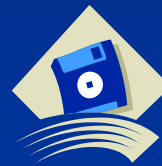
Understand Page Dynamics



Unexpected Unimportant Content

Studying Diff-IE

- Internal study of Diff-IE (3k people, 1+ months)
- Logging
 - Hash of URLs visited
 - Amount of change when revisited
- Feedback buttons
- Survey
 - Prior to installation
 - After 1 month of use
- Experience interview



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"/>
Blogs 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"/>



Diff-IE Results

- People revisit more w/ Diff-IE

- 1st week: 39.4% of visits are revisits
- 4th week: 45.0% of visits are revisits



- Perception of change increases

- Amount of change seen increases

- 1st week: 21.5% revisits changed, by 6.2%
- 4th week: 32.4% revisits changed, by 9.5%



- Diff-IE is driving visits to changed pages

- It supports people in understanding change

Overview

■ Change on the Web

- Content changes over time
- User interaction varies over time (queries, re-visitation, anchor text, query-click stream, “likes”)
- Tools for understanding Web change (e.g., Diff-IE)

■ Improving Web retrieval using dynamics

- Query trends over time
- Retrieval models that leverage dynamics
- Task evolution over time

Temporal Retrieval Models 1

(content-based)

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

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

Change prior

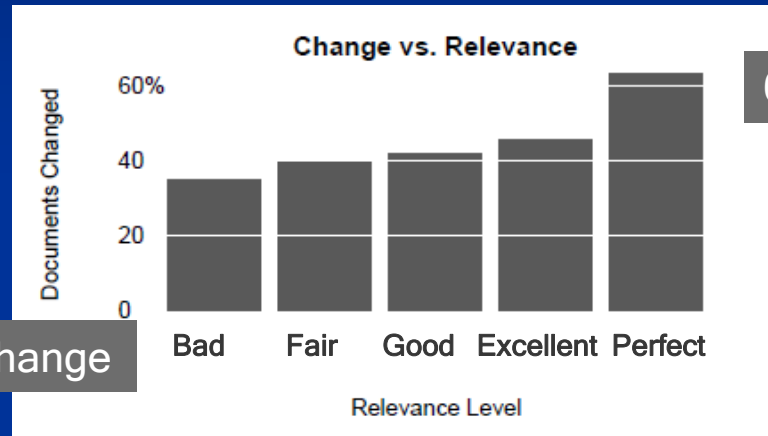
Term longevity

Temporal Retrieval Expt 1

- 18K queries, 2.5M returned documents
 - Crawled docs weekly, for 10 weeks
 - Judged docs for relevance, only once
 - 5-point scale: Perfect/Excellent/Good/Fair/Bad
- Subset of 2k “navigational” queries
 - Queries that have a “Perfect” judgment
 - Assume these relevance judgments are consistent over time
- Measure changes in nDCG

Relevance and Page Change

- Page change is related to relevance



30% “Bad” pages change

60% “Perfect” pages change

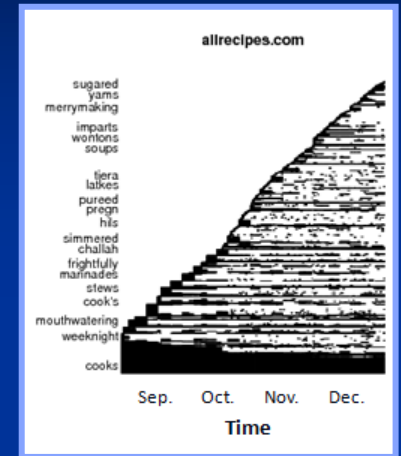
- Use change rate as a document prior (vs. priors based on link structure like Page Rank)
 - Shingle prints to measure change

$$P(D | Q) \propto \boxed{P(D)} \cdot P(Q | D)$$

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

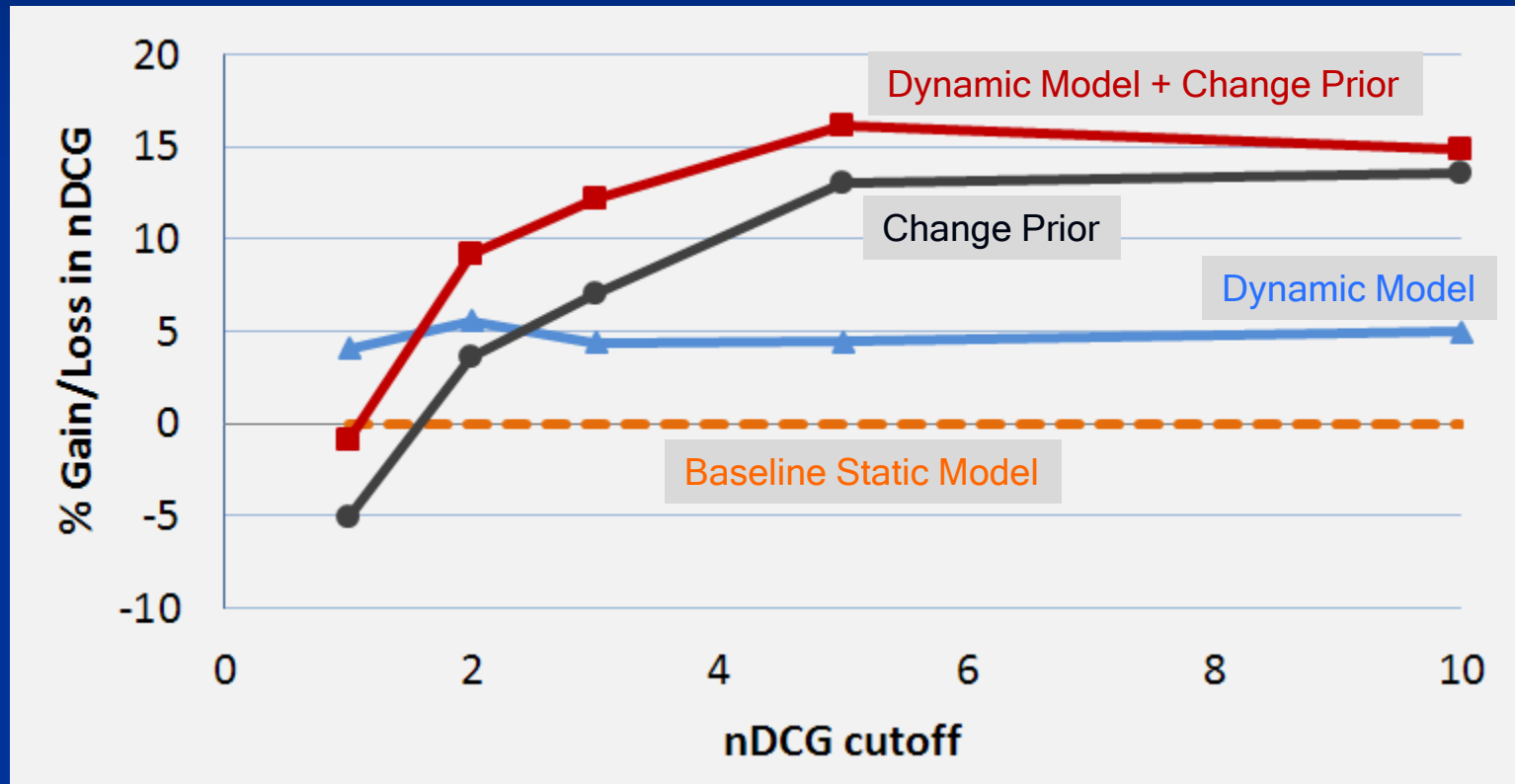


$$P(Q | D) = \lambda_L P(Q | D_L) + \lambda_M P(Q | D_M) + \lambda_S P(Q | D_S)$$

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

↑
Term longevity

Temporal Retrieval Results



Temporal Retrieval Models 2

(behavior-based)

- Initial evaluation: assumed relevance is “static” over time
- But, relevance often changes over time
 - E.g., *SIAM SDM* -- in 2012 vs. in 2011
 - E.g., *US Open 2012* -- in June (golf) vs. in Sept (tennis)
 - E.g., *March Madness 2012* -- before/during/after event
 - Before event: Schedule and tickets, e.g., stubhub
 - During event: Real-time scores, e.g., espn, cbssports
 - After event: General sites, e.g., wikipedia, ncaa
- Current evaluation
 - Collect relevance judgments, query frequency, interaction data, and page content over time

Relevance over Time

- Query: *sigir*
- Why is old content ranked higher?
 - User interaction data (e.g., query-clicks, anchor text) more prevalent for older documents
- Need to weight user behavior signals appropriately

Web Images Videos Shopping News Maps More MSN Home

bing MS Beta 0

Web More▼

sigir

1-10 of 167,000 results · All

ALL RESULTS

Welcome to SIGIR Home
An Iraqi fisherman pushes his boat off-shore to depart on his daily fishing trip. View the Report.
[www.sigir.mil](#) · Mark as spam

ACM SIGIR Special Interest Group on Information Retrieval Home Page
Welcome to the ACM SIGIR Web site. ACM SIGIR addresses issues ranging from theory to user demands in the application of computers to the acquisition, organization ...
[www.sigir.org](#) · Mark as spam

2010

home [ACM SIGIR 2010]
ACM-SIGIR 2010 was held at UniMail, Geneva, Switzerland between 19th
Thanks to all the participants!!! The story continues with ACM-SIGIR 2011.
[www.sigir2010.org](#) · Mark as spam

2009

SIGIR 2009 Archive | SIGIR'09
The SIGIR 2009 conference ran July 19-23, 2009, in Boston, Massachusetts
Hotel and Northeastern University. The conference was chock full of ...
[sigir2009.org](#) · Mark as spam

2011

Welcome to The 34th Annual ACM SIGIR Conference
The 34 th Annual ACM SIGIR Conference Important Dates. 17 Jan 2011 :
papers due; 24 Jan 2011 : Full research paper submissions due
[sigir2011.org](#) · Mark as spam

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 ...
[www.sigir2008.org](#) · Mark as spam

About SIGIR
About SIGIR The Office of the Special Inspector General for Iraq Reconstruction (SIGIR) is the successor to the Coalition Provisional Authority Office of ...
[www.sigir.mil/about/index.html](#) · Mark as spam

ACM SIGIR - News
SIGIR News 2011. Get ready for SIGIR 2011 in Beijing, China! SIGIR invites applications for Donald B. Crouch Travel Grants to help cover the cost of travel, living ...
[www.sigir.org/news.html](#) · Mark as spam

SIGIR - Wikipedia, the free encyclopedia
SIGIR may refer to: Special Inspector General for Iraq Reconstruction; Special Interest Group on Information Retrieval, a Special Interest Group (SIG) of the Association for ...
[en.wikipedia.org/wiki/SIGIR](#) · Mark as spam

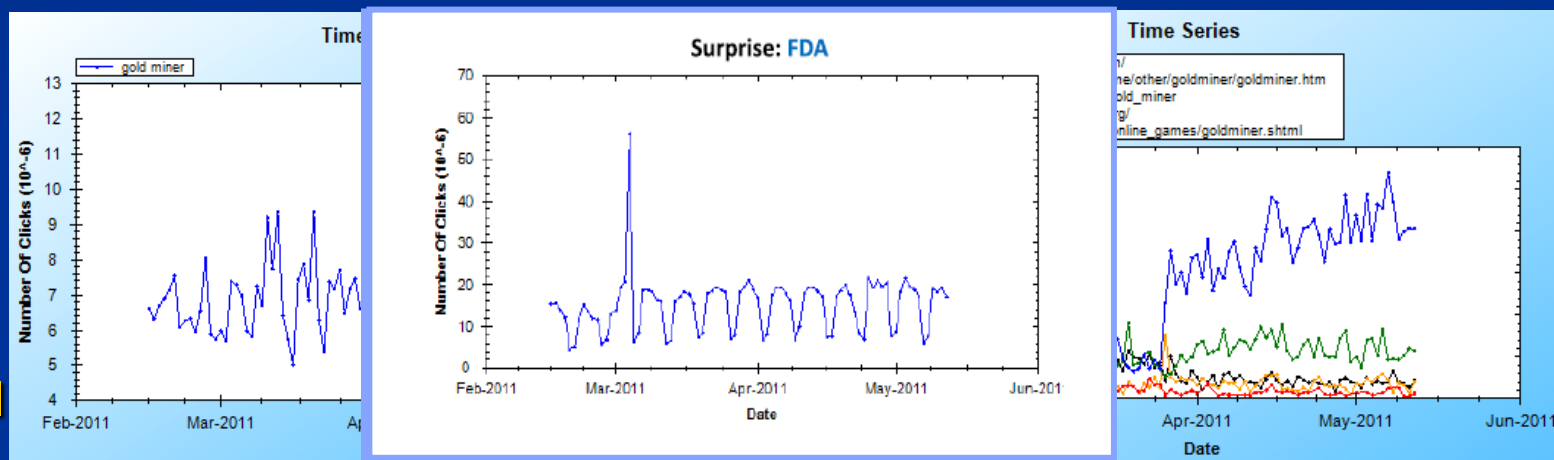
RELATED SEARCHES
SIGIR 2010
SIGIR Iraq
SIGIR Reports
Special Inspector General for Iraq
Special Interest Group on Information Retrieval
Sigir
KDD 2010
ACM SIGIR

SEARCH HISTORY
Search more to see your history
See all
Clear all · Turn off

Temporal Retrieval Expt 2

■ Data

- Queries and clicked URLs, over 4 months



- Actual user search behavior over time (implicit measure)
- Model temporal dynamics of behavior
- Use model predictions to improve ranking

Time Series Modeling

■ Model search behavior as time series

- Assume that the series of behavioral observations $Y_1 \dots Y_n$ is generated sequentially based on some underlying structure (e.g., a sequence of *state vectors*)

■ Linear State Space Model (SSM)

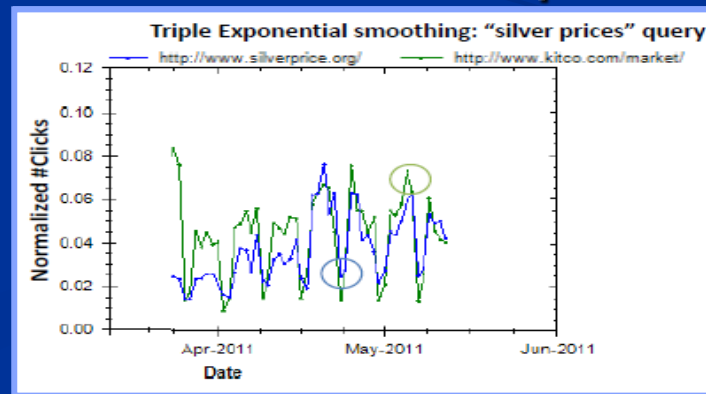
- Let X_t be a state vector at time t , then a semi-linear state space model is defined by:

$$Y_t = W(\theta)X_t + \epsilon_t \quad (\text{observation eqn.})$$

$$X_t = F(\theta)X_{t-1} + G(\theta)\epsilon_{t-1} \quad (\text{state transition eqn.})$$

■ Model state with Holt-Winters decomposition

- Smoothing
- Trend
- Periodic/Seasonal



Experimental Details

- Learn: Time series models of user behavior
 - Can be query or URL dependent
- Predict: Future query and click behavior
- Ranking models
 - Predicted clicks as the only feature for ranking
 - Temporal (+other) features as input to learned ranker
- Three types of features
 - No user behavior (i.e., just content)
 - Historical average of user behavior
 - Uniform, Linear, Power
 - Temporal models of user behavior
 - Smoothing, +Trend, +Trend+Periodicity
- Measure: Correlation (predicted vs. actual) rankings; Win/Loss

Experimental Results

■ Predicted clicks as the only feature

Query Type	Baselines		
	Average	Linear weight	Power weight
General	0.91	0.92	0.93
Tail	0.18	0.21	0.22
Periodic	0.91	0.92	0.93
Dynamic	0.28	0.35	0.38
Alternating	0.80	0.82	0.84
Temp Reform	0.95	0.95	0.95

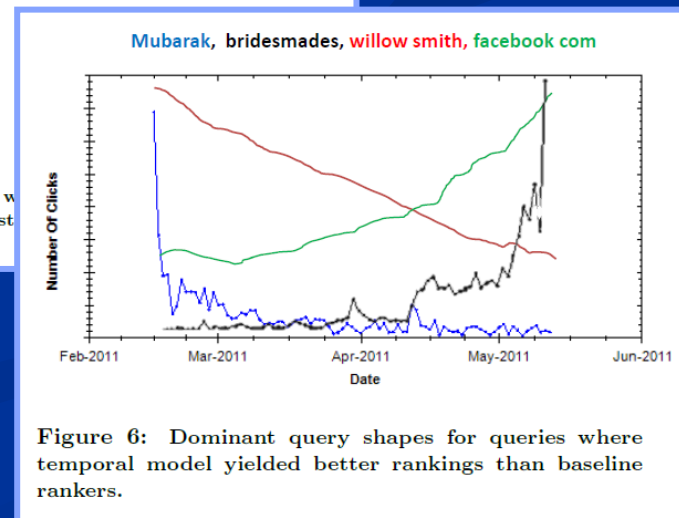
Table 2: Pearson correlation on ordering of our temporal models compared to baseline models. Statistically significant differences based on a paired t-test ($p < .05$) are shown in bold.

■ Ranker trained with content + temporal features

Query Type	No User Behavior	Baseline Models		
	Base Features	Base Features +Average	Base Features +Linear weight	Base Features +Power weight
General	0.47	0.97	0.98	0.98
Tail	0.31	0.20	0.07	0.02
Periodic	0.78	0.87	0.91	0.91
Dynamic	-0.08	0.30	0.30	0.39
Alternating	0.23	0.64	0.90	0.74
Temp Reform	0.19	0.73	0.97	0.96

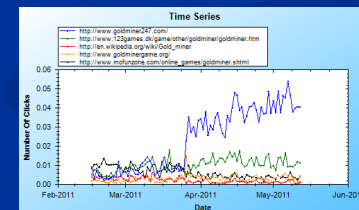
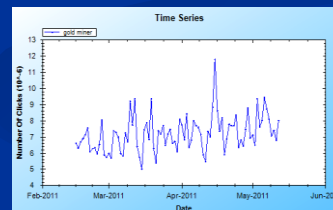
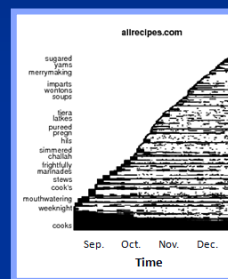
Table 4: Pearson Correlation on ranking using Base features without user behavior, using our temporal models. Statistically significant differences based on a paired t-test performing algorithm ($p < .05$) are shown in bold.

■ Best-performing queries



Temporal IR Summary

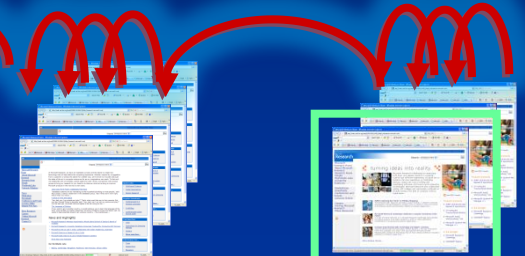
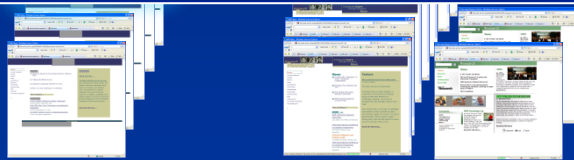
- Improve Web search by modeling temporal dynamics
- Content-based models
 - Rate of page change
 - Detailed term-level changes
- Behavior-based models
 - Query frequency over time
 - URL click patterns over time
- Ongoing work
 - Combing content and behavior features
 - Detecting surprise and periodicity
 - Modeling events



Summary

Temporal IR:
Leverages change
for improved IR

Web content changes: page-level, term-level

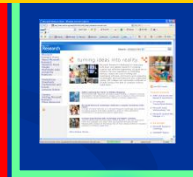
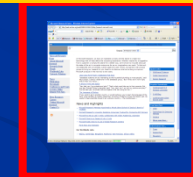


Relating revisitation and change allows us to

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

2006 2007 2008 2009

2006 2007 2008 2009



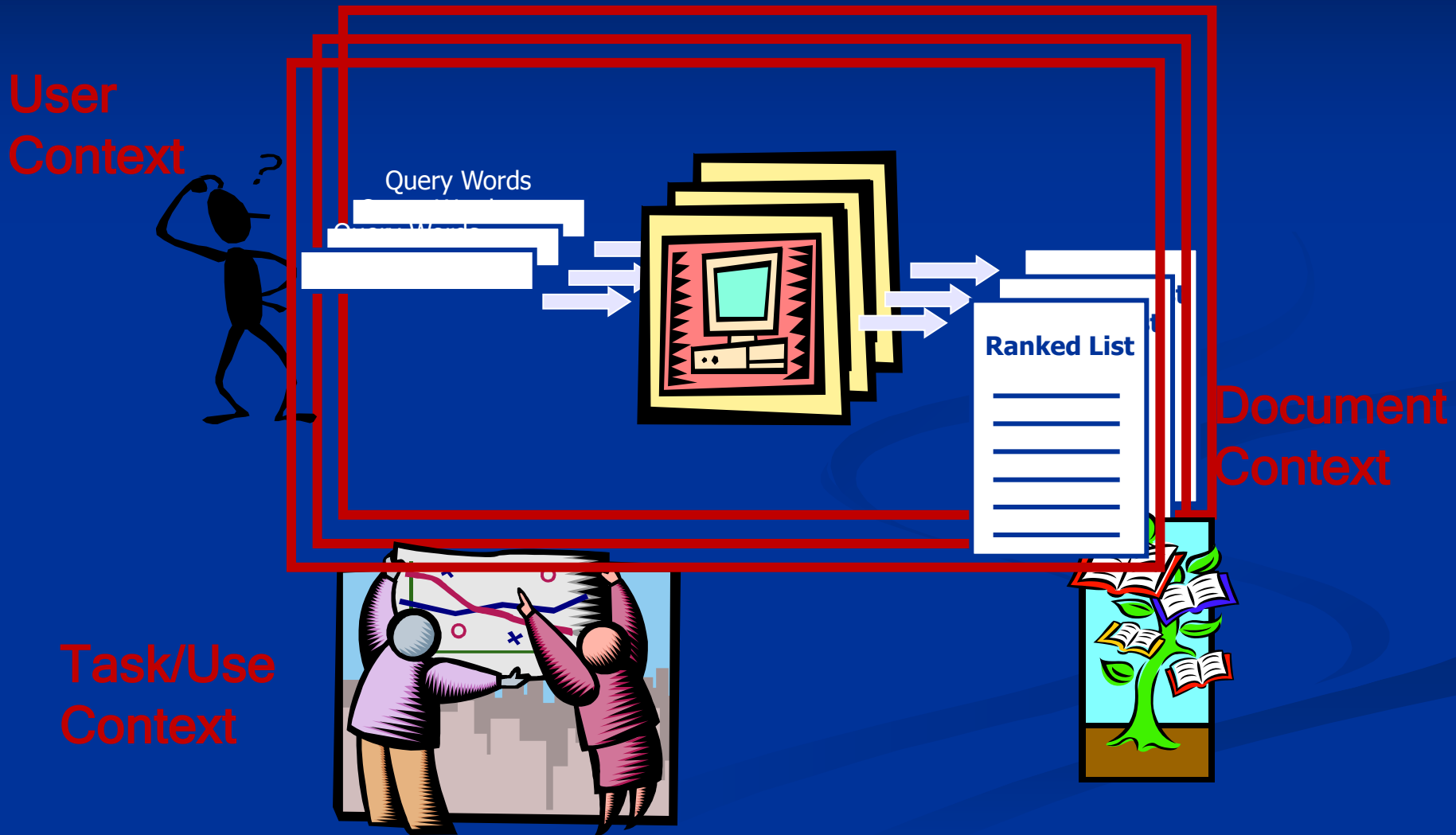
People revisit and re-find Web content

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

Opportunities and Challenges

- 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 for IR
 - Others include: location, individual, tasks, etc.

Think Outside (Search Research) Boxes



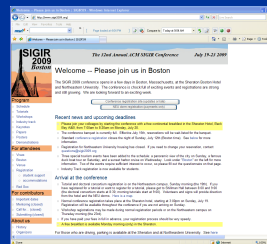
Thank You !

- Questions/Comments ...

- More info,

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

Diff-IE ... try it!



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