

Temporal Dynamics and Information Retrieval

Susan Dumais

Microsoft Research

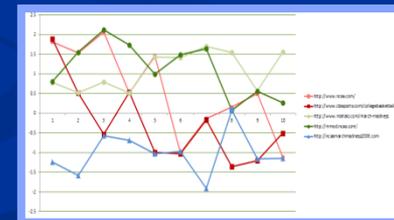
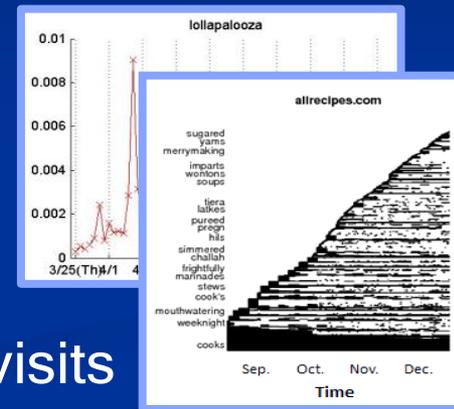
<http://research.microsoft.com/~sdumais>

In collaboration with:

Jaime Teevan, Eytan Adar, Jon Elsas, Dan Liebling,
Richard Hughes, Krysta Svore, Anagha Kulkarni

Change is Everywhere in IR

- New documents (and queries) appear all the time
- Query volume changes over time
- Document content changes over time
- User interaction changes over time
 - E.g., anchor text, query-click streams, page visits
- What's relevant to a query changes over time
 - E.g., *Hurricane Earl* (in Sept 2010 vs. before/after)
 - E.g., *U.S. Open 2010* (in May vs. Sept)
- Change is pervasive in IR systems
 - ... yet, we're not doing much about it !

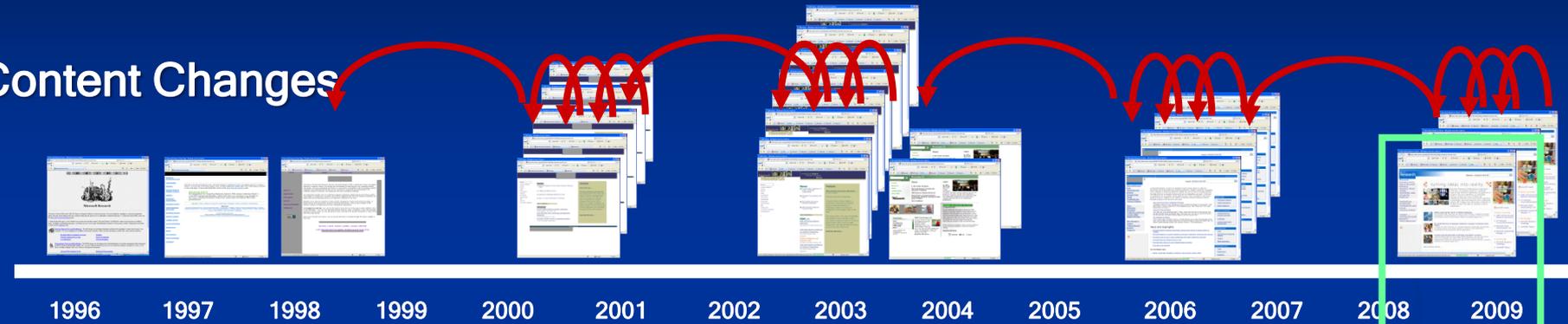


Overview

- Characterize change in digital content
 - Content changes over time
 - People re-visit and re-find over time
 - Relationships between change and re-access
- Improve retrieval and understanding
 - Building tools for understanding change
 - Building IR models that leverage dynamics
- Examples from our work on Web search and browser support ... but more general

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



Information Dynamics

- Characterizing change
 - Content changes over time
 - People re-visit and re-find
 - Relationships between change and re-access
- Improving retrieval and understanding
 - Building support for understanding change (e.g., Diff-IE)
 - Building models and systems that can leverage dynamics (e.g., temporal IR models)

Characterizing 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.

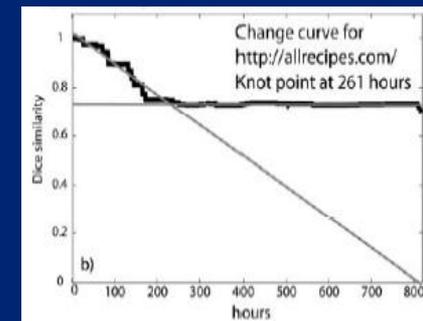
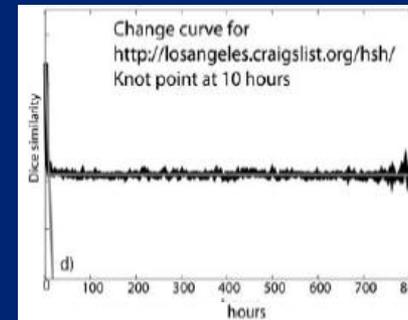
- Amount of change

- Time between changes

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

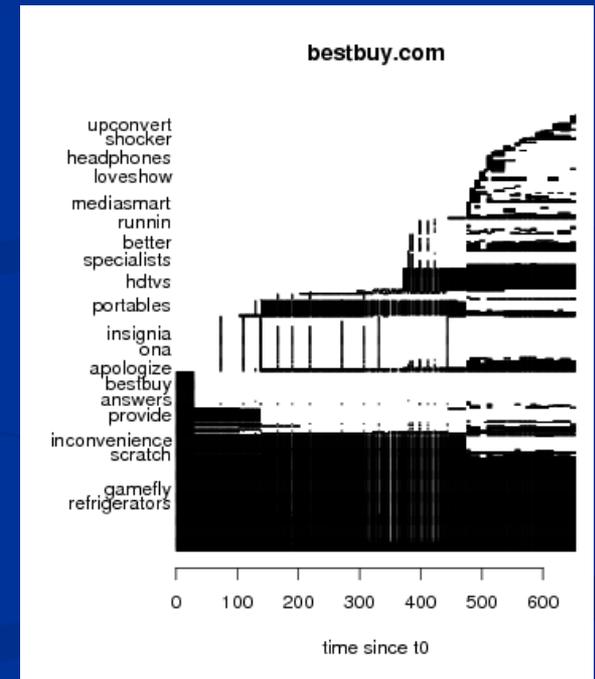
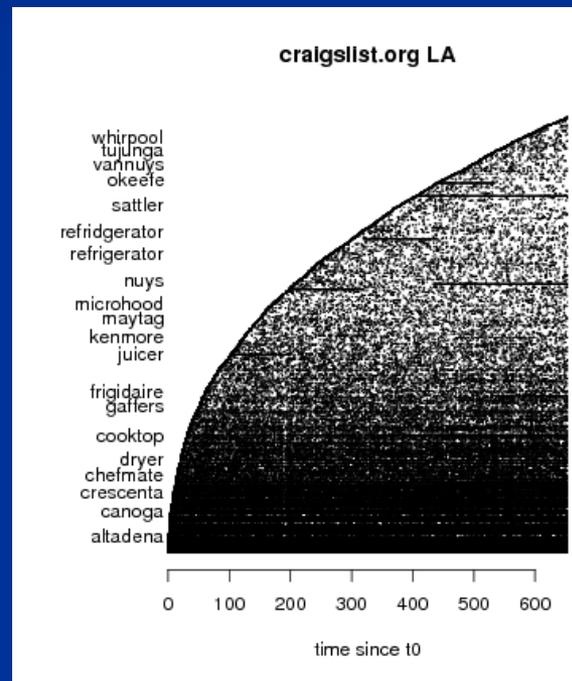
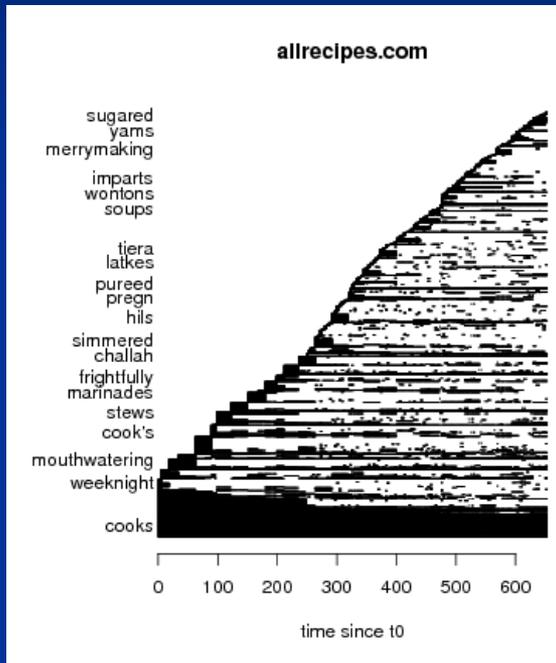


Measuring Within-Page Change

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

The image shows a screenshot of the allrecipes.com website. The website header includes the logo, navigation links (Recipes, Menus, Tips & Advice, NEW! Community, Shop), and a search bar. The main content area features a 'Bake Sale Favorites' section with a photo of breads, a 'Recipe of the Week' for 'Blue Cheese Burgers', and a 'THIS WEEK' section for 'Homemade Tomato Sauce'. A 'Our Community' section features a profile for 'Hills'. At the bottom of the screenshot, a calendar for the month of September is visible, with the word 'Time' centered below the month names (Sep., Oct., Nov., Dec.).

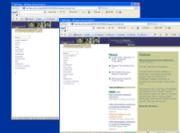
Example Term Longevity Graphs



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

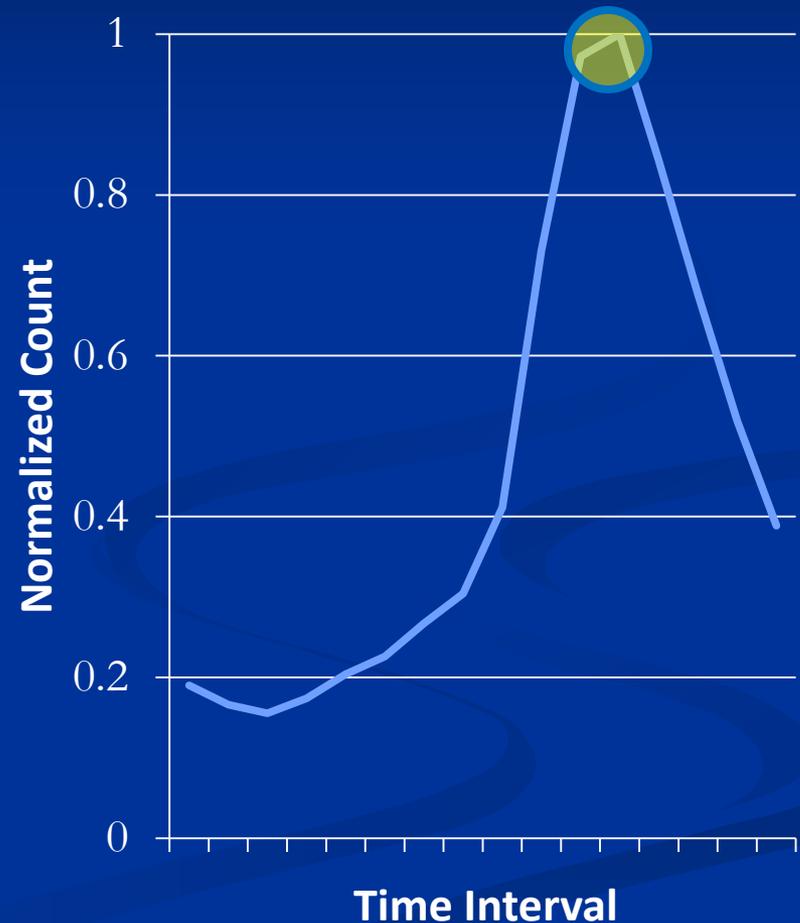


User Visitation/ReVisitation

What was the last Web page you visited?

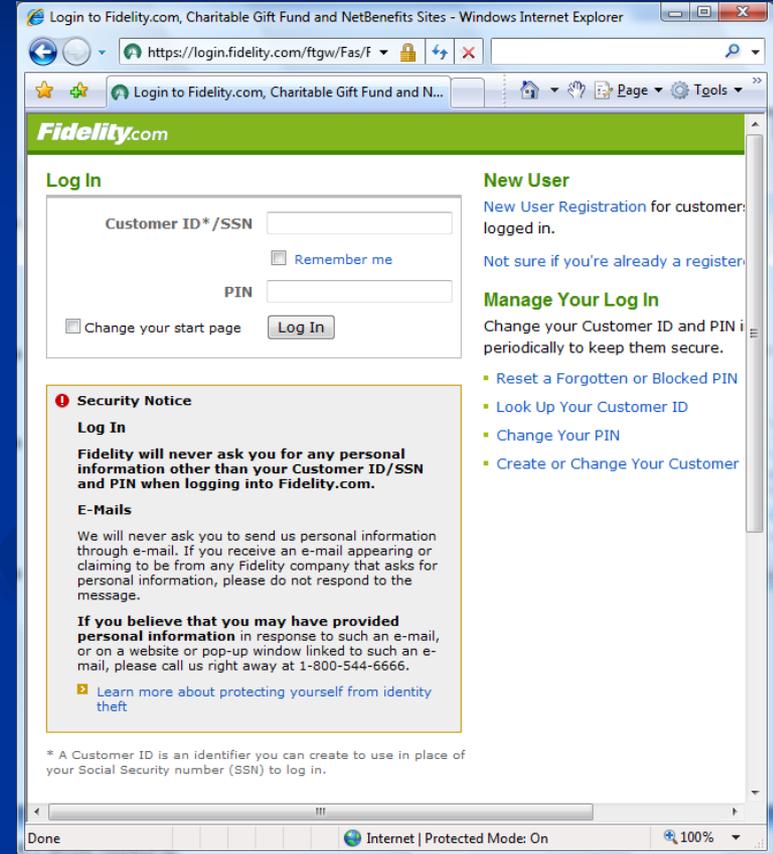
Measuring Revisitation

- 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

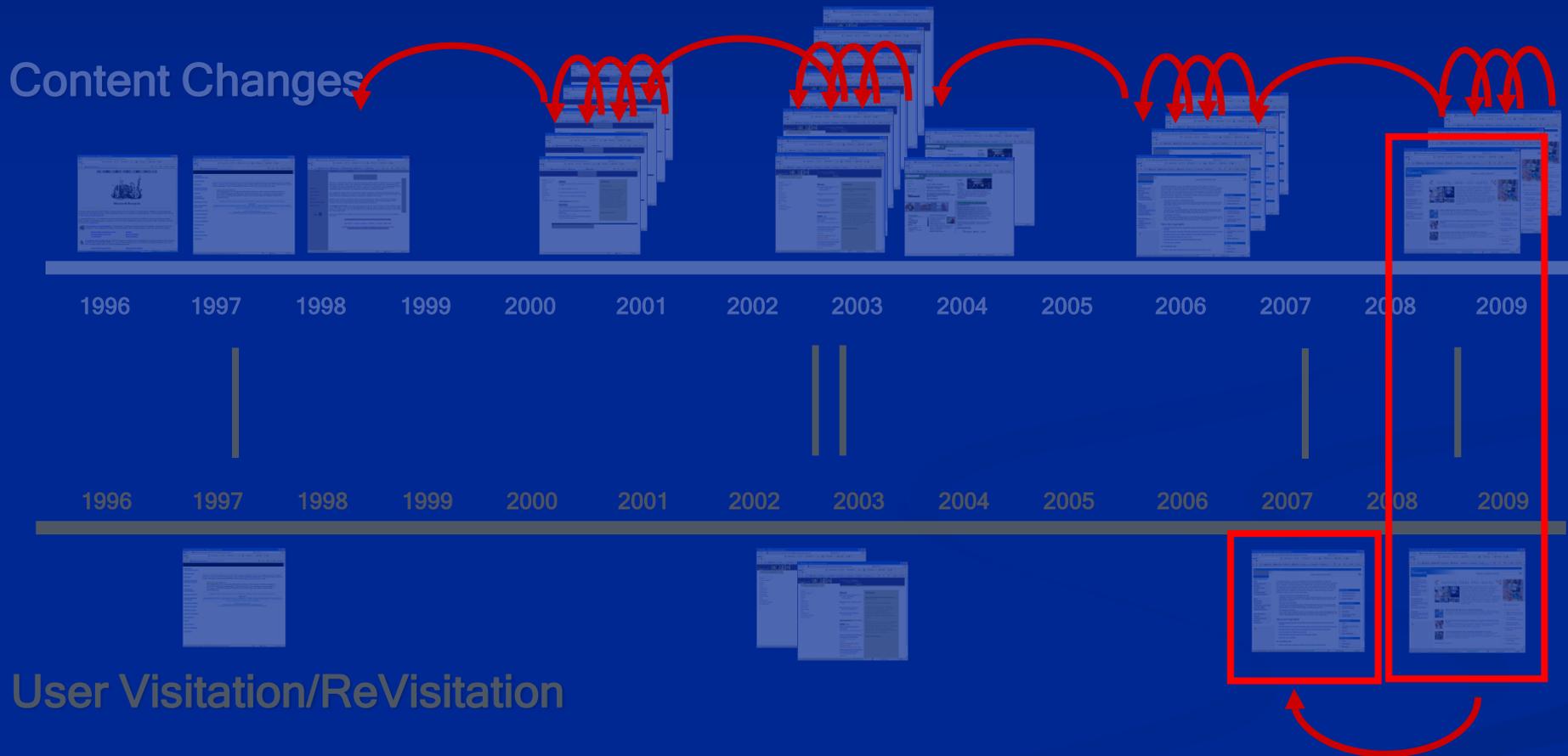


Revisitation and Search (ReFinding)

- Repeat query (33%)
 - Q: *microsoft research*
- Repeat click (39%)
 - <http://research.microsoft.com>
 - Q: *microsoft research, msr ...*
- Big opportunity (43%)
 - 24% “navigational revisits”

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

Relationships Between Revisitation and Change



Why did you revisit the last Web page you revisited?

Possible Relationships Between Change and Revisitation



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

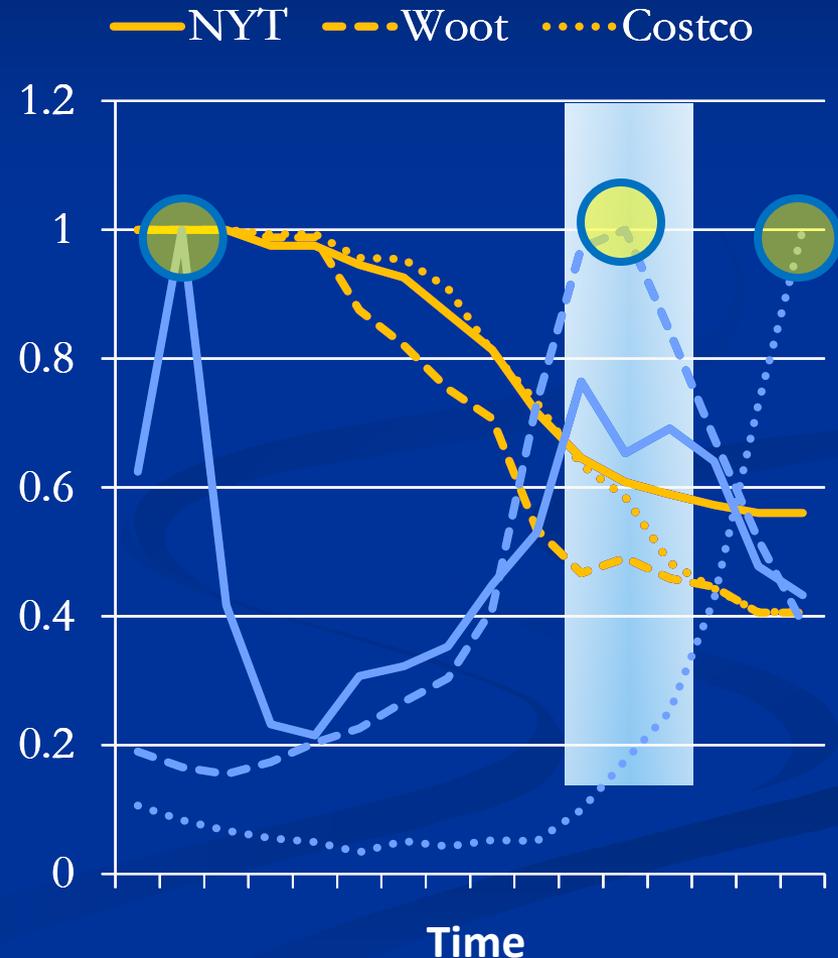
Understanding the Relationship

- Compare summary metrics
 - Revisits: Unique visitors, visits/user, interval
 - Change: Number of, interval between, Dice coeff.

	Number of changes	Time between changes	Dice coefficient
2 visits/user	172.91	133.26	0.82
3 visits/user	200.51	119.24	0.82
4 visits/user	234.32	109.59	0.81
5-6 visits/user	269.63	94.54	0.82
7+ visits/user	341.43	81.80	0.81

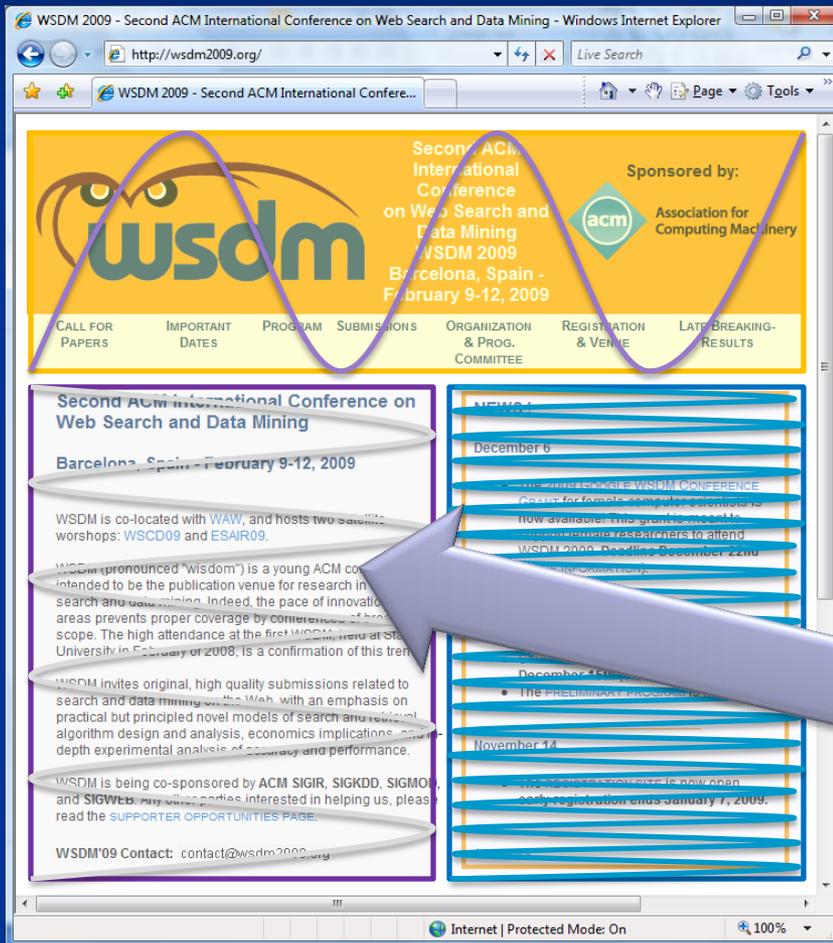
Comparing Change and Revisit Curves

- Three pages
 - New York Times
 - Woot.com
 - Costco
- Similar change patterns
- Different revisitation
 - NYT: *Fast* (news, forums)
 - Woot: *Medium*
 - Costco: *Slow* (retail)



Within-Page Relationship

- Page elements change at different rates
- Pages are revisited at different rates



- “*Resonance*” can serve as a filter for identifying interesting content

Protect and Provide.
HSBC Insurance

The New York Times

Thursday, August 7, 2008 Last Update: 3:02 PM ET

Instant* Term Life.
HSBC Insurance

 Search

Get Home Delivery | Personalize Your Weather

- JOBS
- REAL ESTATE
- AUTOS
- ALL CLASSIFIEDS
- WORLD
- U.S.
- Politics
- Washington
- Education
- N.Y./REGION
- BUSINESS
- TECHNOLOGY
- SPORTS
- OLYMPICS
- SCIENCE
- HEALTH
- OPINION
- ARTS
- Books
- Movies
- Music
- Television
- Theater
- STYLE
- Dining & Wine
- Fashion & Style
- Home & Garden
- Weddings/ Celebrations
- TRAVEL
- Blogs
- Cartoons / Humor
- Classifieds
- Corrections
- Crossword/ Games
- First Look
- Learning Network
- Multimedia
- NYC Guide
- Obituaries
- Podcasts

Coalition in Pakistan Moves to Impeach Musharraf

By JANE PERLEZ 12:19 PM ET
 Pakistan's usually fractious coalition government moved decisively for the first time on Thursday to impeach President Pervez Musharraf.

• Times Topics: Pervez Musharraf | Pakistan

[Post a Comment](#) | [Read \(33\)](#)

E.P.A. Won't Ease Requirements for Ethanol in Gas

By MATTHEW L. WALD 23 minutes ago

The agency denied a request to cut the quota for the use of ethanol in cars, concluding that the goal of reducing the nation's reliance on oil trumps any effect on food prices.

In Anthrax Scientist's E-Mail, Hints of Delusions

By ERIC LIPTON

The e-mail messages of Bruce E. Ivins provide a narrative of his paranoid episodes, dating to 2000.

- The Lede: Reactions to the Case 12:22 PM ET
- F.B.I. Presents Anthrax Case, Saying Scientist Acted Alone
- Times Topics: Anthrax

2008 BEIJING OLYMPICS



Elizabeth Dalziel/Associated Press

SPORTS OF THE TIMES

Raising the Bar and Raising Elbows

By GEORGE VECSEY 11:07 AM ET

Increased physicality in women's basketball will not give a black eye to the entire sport. However, it most certainly gave a black eye to Penny Taylor of Australia on Tuesday.

INTERACTIVE SCHEDULE

Friday	Saturday	Sunday
1 events	4 events	18 events
1 event	1 event	23 events

⚡ Olympics Tracker

Follow the schedules of your favorite sports, day by day and event by event.

Judge Orders Detroit Mayor Jailed

By NICK BUNKLEY 39 minutes ago

OPINION »

An Olive Branch From the Dalai Lama

By NICHOLAS D. KRISTOF

For the first time, the Tibetan leader has expressed a willingness to accept Communist Party rule over his homeland.



[Post a Comment](#) | [Read \(104\)](#)

THEATER »

With Echoes of Chekhov

A. R. Gurney's "Buffalo Gal" is about an actress who retreads the boards of her hometown.



• Times Topics: A. R. Gurney

OPINION »

- Collins: Energy Drill
 - Cohen: Spanish Summer
 - Egan: Nanny Nation
 - Self: Vanishing Cliffs
 - Editorial: Hamdan's Trial
- [Comments \(221\)](#)

AMERITRADE
\$9.99 trades. [GO](#)
 No guesswork.

MARKETS » AT 3:05 PM ET

S.&P.	Dow	Nasdaq
500	11,481.02	2,362.26
1,271.12	-175.05	-16.11
-18.07	-1.50%	-0.68%
-1.40%		

GET QUOTES My Portfolios »

Stock, ETFs, Funds [Go](#)

With every purchase, Visa prevents, detects and resolves online fraud.

SAFE. SECURE.



Go to Visa.com/Secure for details



Hi. Are you new? [Start here.](#)

[GO!](#)
 username password ([I forgot!](#))

HP Compaq 6820s 2GHz 17" Widescreen Notebook

\$529.99

+ \$5 shipping

CONDITION: **New**

PRODUCT: 1 HP Compaq 6820s Celeron 2.0GHz, 160GB, 17" Widescreen, 2GB DDR2, DL-DVDRW, 802.11a/b/g, Vista

[DISCUSS THIS PRODUCT \(176 COMMENTS\)](#)



[PRODUCT INFO](#) [PRODUCT STATS](#)

Sometimes The Past Is Hard To Let Go

"-and when she told me, I just couldn't deal. So I left, just like I always did. Just like my father."

Tom, thank you for sharing all that. Have one of these tissues. Do you feel like you've made a forward step?

"I do, Jack, I really do. Can I maybe have a second tissue?"

As many as you like, Tom, you don't even need to ask. Let's hear from someone else now. HP Compaq 2GHz Celeron 17" Widescreen Notebook, tell us more about your loss.

Well, with me, it was the teasing. See, I'm in the tech industry, which is a really high stress field. And, you know, growing up, everyone in a position of authority says, you know, play the hand you're dealt. But I've got a 2.0GHz Intel Celeron 550 Processor with 1MB L2 cache and 533MHz front side bus, and most people think that's nothing. But, you know, what are they doing? Playing Bejeweled on their lunch break and writing out memos in work? Maybe looking for um... nee... oh... are... en?

WOOTCAST

"Listener Mailbag" (4 comments)

THIS WEEK ON **wine.woot!**

AUG 07 Laura Zahtila Vineyards Napa Valley Chardonnay - Three Pack

TODAY ON **shirt.woot!**

AUG 07 Sweet Dreams Are Made of These...

SONY

The new Sony® VAIO® FW notebook with extra-wide screen.



High Definition. It's in our DNA.

LOGIN

Search in All **SHOP FOR BUSINESS PRODUCTS... CLICK BUSINESS ABOVE**

OFFERS ENDING SUNDAY

fall savings COSTCO.COM

NEW OFFERS

CLICK HERE

CLICK HERE



↖ Select Dell Desktops & Laptops
Up to \$400 OFF



↖ Q-See 16 CH Security System
\$999.99 After \$1,000 OFF



TOSHIBA SHARP

↖ Select Branded TV Hot Buys
Up to \$400 OFF

USA

Sign up to receive offers by e-mail:

your e-mail address

Earn 3% cash back when you buy gas.



THE TALES OF BEEDLE THE BARD

\$7.39 DELIVERED

↖ CLICK HERE

THE SNOWBALL: Warren Buffett and the Business of Life

\$19.99

↖ CLICK HERE TO PRE-ORDER

for your baby

↖ CLICK HERE



↖ Sunsetter Motorized Retractable Patio Awnings 30%



↖ Hoover All Terrain SteamVac
\$199.99 After \$50 OFF



↖ Emergency Food Kit
\$74.99 After \$10 OFF



↖ 4-pc Hardside Luggage
\$149.99 After \$50 OFF



↖ Garmin nuvi 770 4.3" GPS
\$399.99 After \$100 OFF



↖ Saeco Deluxe Espresso
\$699.99 After \$200 OFF

Online Savings Account

3.75% APY*

on balances of \$10,000 or more

MEMBER FDIC



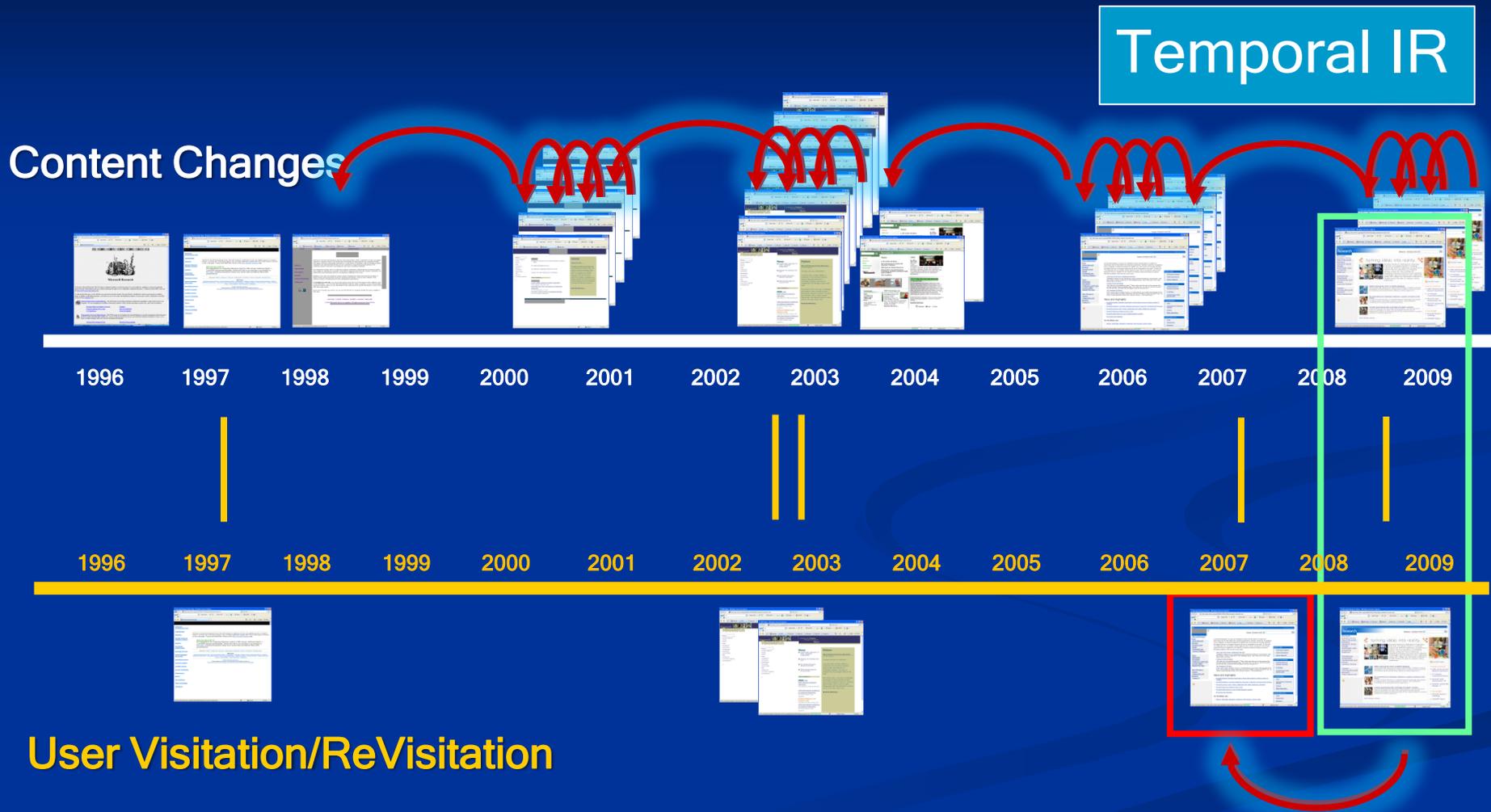
Give us your Best Shot



Information Dynamics

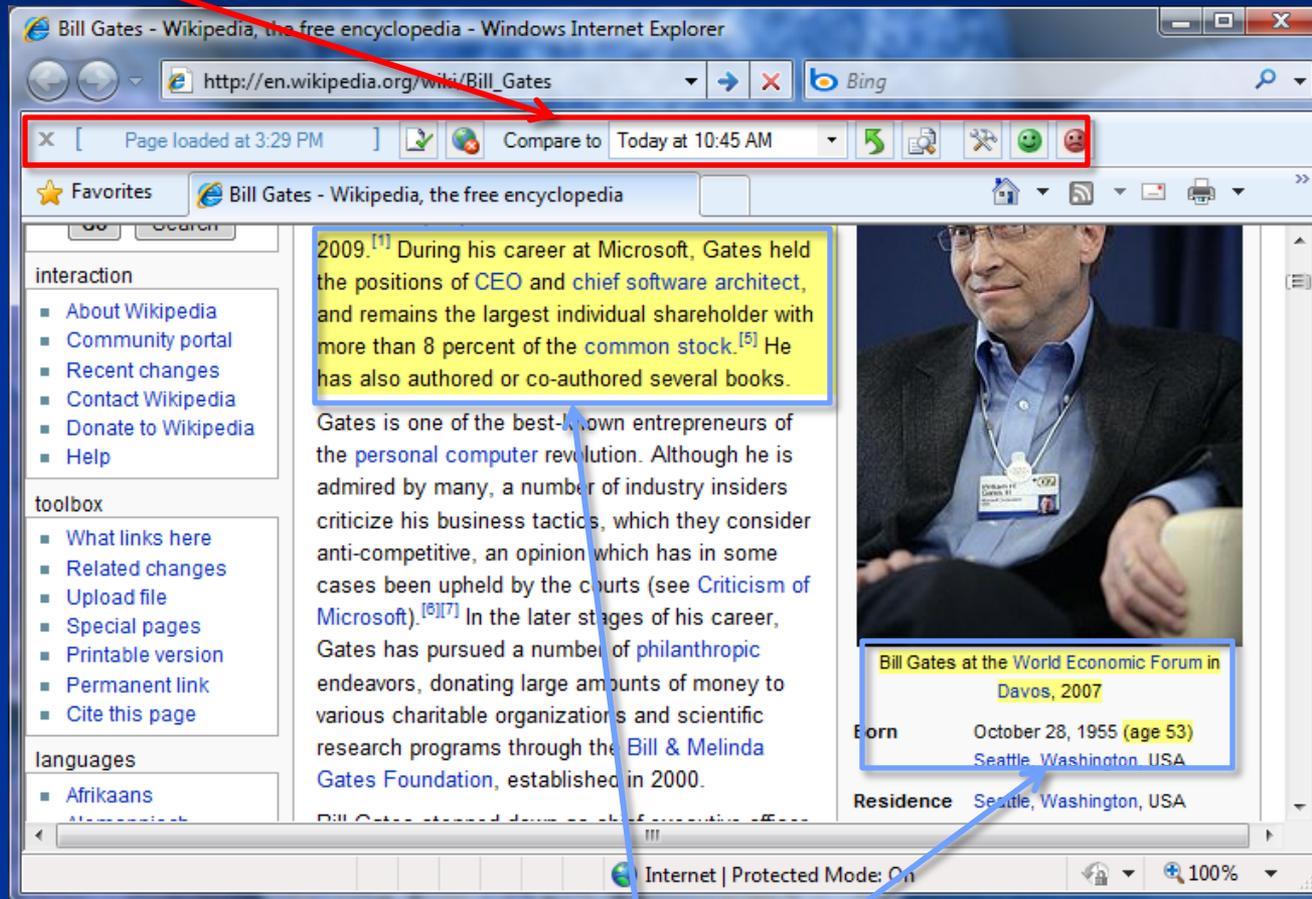
- Characterizing change
 - Content changes over time
 - People re-visit and re-find
 - Relationships between change and re-access
- Improving retrieval and understanding
 - Building support for understanding change (e.g., Diff-IE)
 - Building models and systems that can leverage dynamics (e.g., temporal IR models)

Building Support for Web Dynamics



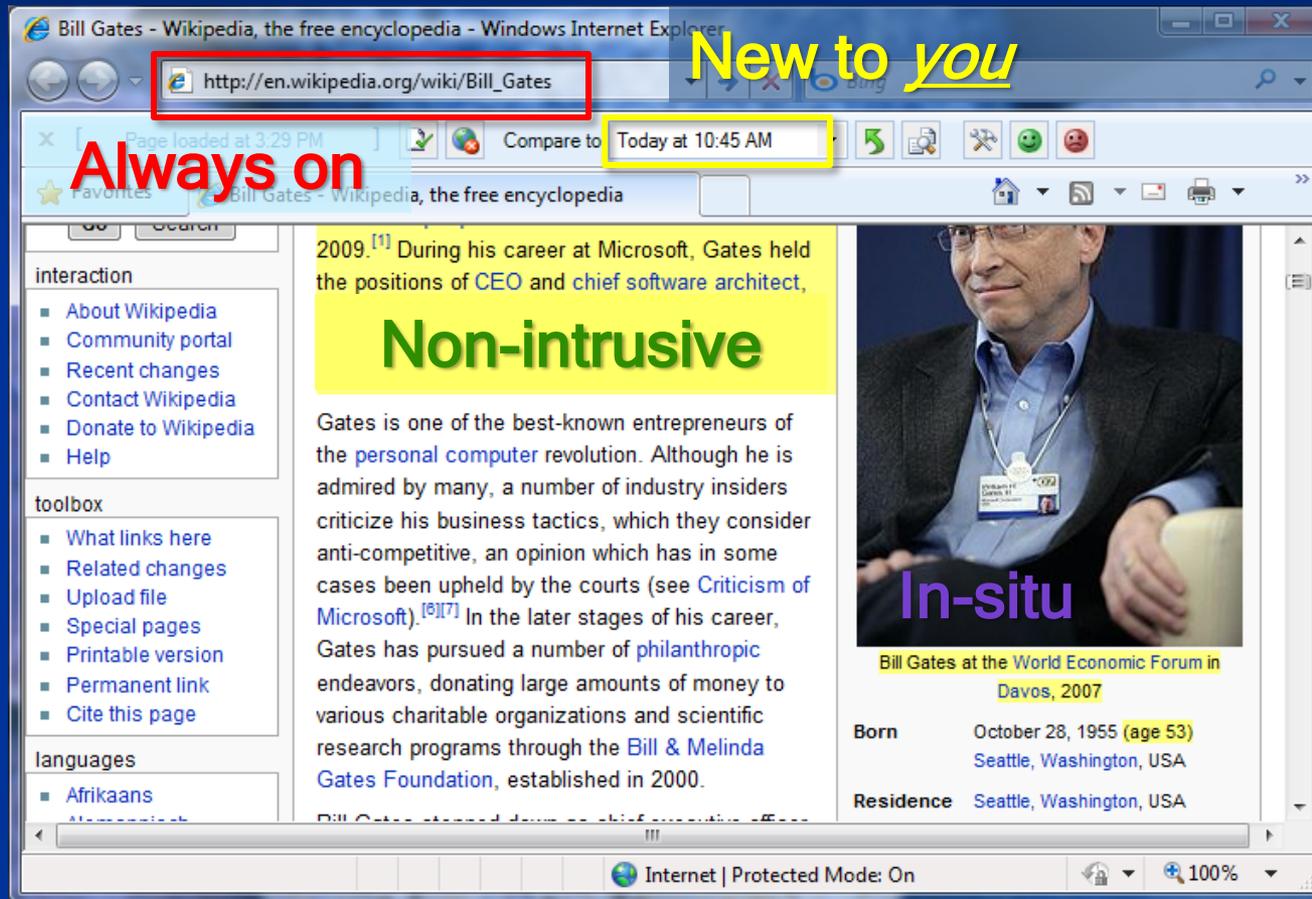
Diff-IE

Diff-IE toolbar



Changes to page since
your last visit

Interesting Features of Diff-IE



Examples of Diff-IE in Action

Expected New Content

msnbc
Powered by Bing
MSN Home | Mail | More | Sign In
featuring TODAY | Nightly News | Dateline | Meet the Press | msnbc tv | NBC Sports

Top stories as of 10:14 a.m. Wed., Sept 30, 2009
New York, NY
mostly cloudy, 57°F

World news

Thousands trapped in 2nd Pacific-area quake, official says

UPDATED At least 75 people dead as 7.6-magnitude temblor hits Indonesia; hundreds of houses reportedly damaged; panicked residents flee their homes. [Full story](#) | [Video: Western Sumatra struck](#)

Iraq resonates in claims over Iran nukes

NYT: While similarities between 2002, with its faulty IWMMD estimates, and 2009 are unmistakable, the differences are profound. [Full story](#)

China stalls action against Iran

Beijing's reluctance stems from oil, ideology. [Story](#)

[Video: World leaders meet on Iran nuclear threat](#)

Damaged boat (John Newton / AFP - Getty Images)

Samoans count cost after deadly tsunami

Undersea earthquake propels waves onto island nation and U.S. territory; toll expected to rise. [Story](#) | [Video: South Seas terror](#) | [Are you there? Send images, reports](#)

Sports

Some in Windy City hope 2016 bid is a bust

"I know it's going to cost u money somehow," one Chicago resident says. [Full story](#)

NEW Meet the Olympic contestants: Chicago, Tokyo, Rio, Madrid

[Video: Obamas and Winfrey make '16 pitch](#)

U.S. news

Guns to be allowed in Arizona bars | Vote

[Video: Summit to tackle distracted drivin](#)

advertisement | your ad here

We do right by you and your money. **ally** Straightforward.

Mouse over tabs for details

9-Month No Penalty CD

Great Rates

Categories

- U.S. news
- World news
- Politics
- Business
- Sports
- Entertainment

Explore

More news | My news | Pinned

Other top stories

- Swine flu school closings could cost billions
- 1st lady in Europe for Olympics vote | [Video](#)
- GOP candidate: Too busy with family to vote
- Newsweek: What most suicide bombers have in common
- UPDATED** Church ordains sex offender as minister

Only on msnbc.com

Reuters

Slideshow: China

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' with statistics: 20 following, 12 followers, and 1 listed. Below this are sections for 'Tweets' (0) and 'Favorites'. The 'Following' section displays a grid of profile pictures for users being followed, including logos for CNN, Huffington Post, and NY Times, as well as individual user avatars. The main content area lists the profiles of the users being followed, each with their name, bio, and a recent tweet. The users listed are pfromthenc, DARPA_News, huffingtonpost, nytimes, and dm Russell.

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	
 dm Russell 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

Social Programme

CALL FOR CONTRIBUTIONS

- Call for Research Papers
- Call for Posters and Demos
- Call for Panels
- Call for Tutorials
- Call for Workshops
- Doctoral Consortium
- Submissions
- Important Dates
- Sponsorship Programme

ORGANISATION

- Programme Committee
- Organising Committee
- Contact
- Sponsors

RELATED CONFERENCES

- ACM/IEEE Joint Conference on Digital Libraries
- International Conference on Asia-Pacific Digital Libraries
- International Conference on Digital Libraries
- Russian Conference on Digital Libraries

All changes will be announced here so please check for updates regularly. Thanks!

Invitation: Excursion to Auchentoshan distillery

SEPTEMBER 2, 2010



On Monday afternoon you are invited to join the tour to Auchentoshan distillery for a discovery of the famous Scottish Single Malt Whisky. This is something not to be missed! Free tickets will be available on a first-come first-served basis at the conference registration desk in Boyd Orr building on Monday from 8am onwards.

The coach to Auchentoshan departs from Hilton City Centre at 3.30pm and from the University at 3.50pm. If you – due to flight times etc – can't make it to the registration desk before the coach leaves please drop us an email to info@ecdl2010.org before Sunday evening. We might then be able to allocate some tickets.

For tutorials and doctoral consortium delegates there will be a later departure. Please refer to the [social programme page](#) for additional information.

Shuttle service to Uni Campus

SEPTEMBER 1, 2010

Delegates staying at either Wolfson Hall or Hilton City Centre can use the following complimentary shuttle service to Uni Campus in the mornings:

Departure	Monday, Sept 6	Tuesday, Sept 7	Wednesday, Sept 8	Thursday, Sept 9	Friday, Sept 10

ECDL2010 on Twitter

 [hvdsomp](#) Bumped by ongoing data modeling issues with open annotation. Also because I need to present a tutorial on it, Monday at #ecdl10 4 days ago

 [ECDL2010](#) We are happy to announce a free shuttle service from the main accommodation of #ECDL2010 to Uni campus in the mornings. <http://www.ecdl.org> 4 days ago

 [ECDL2010](#) Explore the world famous Scottish Whisky! #ECDL2010 invites all delegates to Auchentoshan Distillery

twitter Join the conversation

NEED HELP?

Our ECDL volunteers are here to help:



Serendipitous Encounters



CIKM 2010
19th ACM International Conference on Information and Knowledge Management

Toronto, Canada
October 26 - 30, 2010

Participants | Program | Contributors | Locations | About Us | Search

Updated News

Best Paper Award Shortlist Posted
October 24, 2010
CIKM'10 Best Inter-disciplinary Paper Award and Best Student Paper Award Shortlist is posted here.

Final Program Posted
October 22, 2010
CIKM'10 Final program is posted here.

Student Travel Awards Posted
October 20, 2010
The list of student travel awards is posted here.

Poster Guidelines Posted
October 7, 2010
Instructions on the size of posters for the main conference and workshops are posted here.

Industry Event Information Updated
October 1, 2010
The CIKM'10 industry event information is updated here.

Platinum Sponsors
Microsoft
Research

Gold Sponsors
eBay

SOSO 搜搜

Understand Page Dynamics

The screenshot shows a Bing search results page for the query "cikm 2010". The search bar at the top contains the text "cikm 2010" and a magnifying glass icon. Below the search bar, there are tabs for "Web", "Images", and "More". The search results are displayed in a list format. The first result is titled "The 19th ACM International Conference on Information and Knowledge ..." and includes a link to "www.yorku.ca/~cikm10". The second result is titled "CIKM'10" and includes a link to "www.yorku.ca/cikm10/papers.php". The third result is titled "CIKM 2010 : International Conference on Information and Knowledge ..." and includes a link to "www.wikicfp.com/cfp/servelet/event.showcfp?eventid=5822©ownerid=2". The fourth result is titled "Conference on Information and Knowledge Management (CIKM)" and includes a link to "cikmconference.org". The fifth result is titled "CIKM2010 at Pipers (inside Fairmont Royal York Hotel) (Tuesday ...)" and includes a link to "upcoming.yahoo.com/event/6623508/ON/Toronto/CIKM2010/Pipers-inside-Fairmont-Royal-York...". The sixth result is titled "CIKM 2010 — PaIR 2010" and includes a link to "pair.ir-facility.org/links/cikm-2010". The seventh result is titled "SMUC 2010 — 2nd International Workshop on Search and Mining User ..." and includes a link to "labs.brainsins.com/events/smuc2010". At the bottom right of the page, there are links for "Internal preview" and "Help improve Bing".

Web Images Videos Shopping News Maps More | MSN Hotmail Sign in Redmond, Washington Preferences

bing MS Beta br1009 cikm 2010

Web Images More

ALL RESULTS 1-50 of 113,000 results · Advanced

[The 19th ACM International Conference on Information and Knowledge ...](#)
Updated News. Poster Guidelines Posted October 7, 2010 Instructions on the size of posters for the main conference and workshops are posted here. Industry Event Information ...
www.yorku.ca/~cikm10 · Mark as spam

CIKM'10
CIKM'10 – Session Details. CIKM 2010 Accepted Full Papers . Paper ID. Title and Authors . Session 1A ...
www.yorku.ca/cikm10/papers.php · Mark as spam

Show more results from www.yorku.ca

[CIKM 2010 : International Conference on Information and Knowledge ...](#)
CIKM 2010 : International Conference on Information and Knowledge Management | Series: CIKM 2010, CIKM 2011, CIKM 2012
www.wikicfp.com/cfp/servelet/event.showcfp?eventid=5822©ownerid=2 · Mark as spam

[Conference on Information and Knowledge Management \(CIKM\)](#)
CIKM 2010 will be held in Toronto, Canada, October 25-29, 2010 General Chair: Jimmy Huang, University of York, Canada Email: "Jimmy Huang" <jhuang__at__yorku.ca>
cikmconference.org · Mark as spam

[CIKM2010 at Pipers \(inside Fairmont Royal York Hotel\) \(Tuesday ...](#)
Registration (CIKM 2010) Information If you have any questions regarding registration, please send an e-mail to cikm10@yorku.ca Each accepted paper must be accompanied ...
upcoming.yahoo.com/event/6623508/ON/Toronto/CIKM2010/Pipers-inside-Fairmont-Royal-York... · Mark as spam

[CIKM 2010 — PaIR 2010](#)
CIKM 2010 will take place in Toronto, Canada. Toronto is Canada's largest city. More than 60% of the US population live within a 90-minute flight to Toronto.
pair.ir-facility.org/links/cikm-2010 · Mark as spam

[SMUC 2010 — 2nd International Workshop on Search and Mining User ...](#)
SMUC 2010. 2nd International Workshop on Search and Mining User-generated Contents, colocated at CIKM 2010
labs.brainsins.com/events/smuc2010 · Mark as spam

Internal preview Help improve Bing

Expected



Expected New Content



Monitor



Unexpected Important Content

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

Subject	Replies	Views	Last post
Small Business + ...	21	206	10/29/2010 10:00 AM
Coaching Inquiries	10	93	10/29/2010 10:00 AM
Back to work Part II	8	43	10/29/2010 10:00 AM
Aldergate Christian Preschool - HIKING	0	17	10/29/2010 10:00 AM
do you know anyone in Italy or Sweden?	3	20	10/29/2010 10:00 AM
Call an Interview and Hire Recruiting	94	562	10/29/2010 10:00 AM
Autism and Industry? Finding the Right Fit	7	105	10/29/2010 10:00 AM
Any members work at Amazon.com?	10	118	10/29/2010 10:00 AM
Time Magazine on SAHRA - getting back into work	1	75	10/29/2010 10:00 AM
Freemasonry Resurgence	13	41	10/29/2010 10:00 AM
Congrat designer or Admin?	0	45	10/29/2010 10:00 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"/>				
Message boards, forums, newsgroups	<input type="radio"/>				
Company homepages	<input type="radio"/>				
Personal homepages of people you know	<input type="radio"/>				
Pages with product information	<input type="radio"/>				
Reference pages (dictionaries, yellow pages, maps)	<input type="radio"/>				
Wikipedia pages	<input type="radio"/>				
Blips you read	<input type="radio"/>				
Search engine results	<input type="radio"/>				

In situ
Representative

■ Logging

- URLs visited
- Amount of change when revisited



Experience

Longitudinal

■ Experience interview



People Revisit More

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

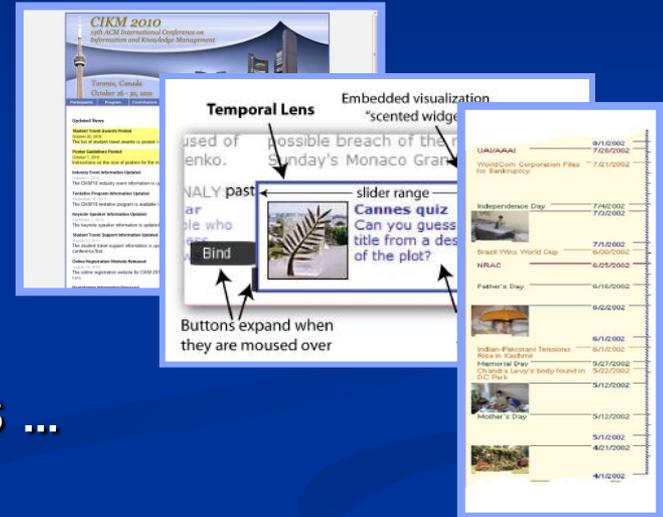
Revisited Pages Change More

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

Other Examples of Dynamics and User Experience

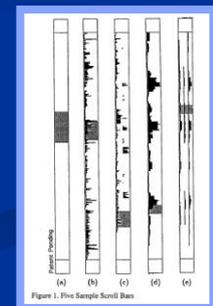
■ Content changes

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



■ Interaction changes

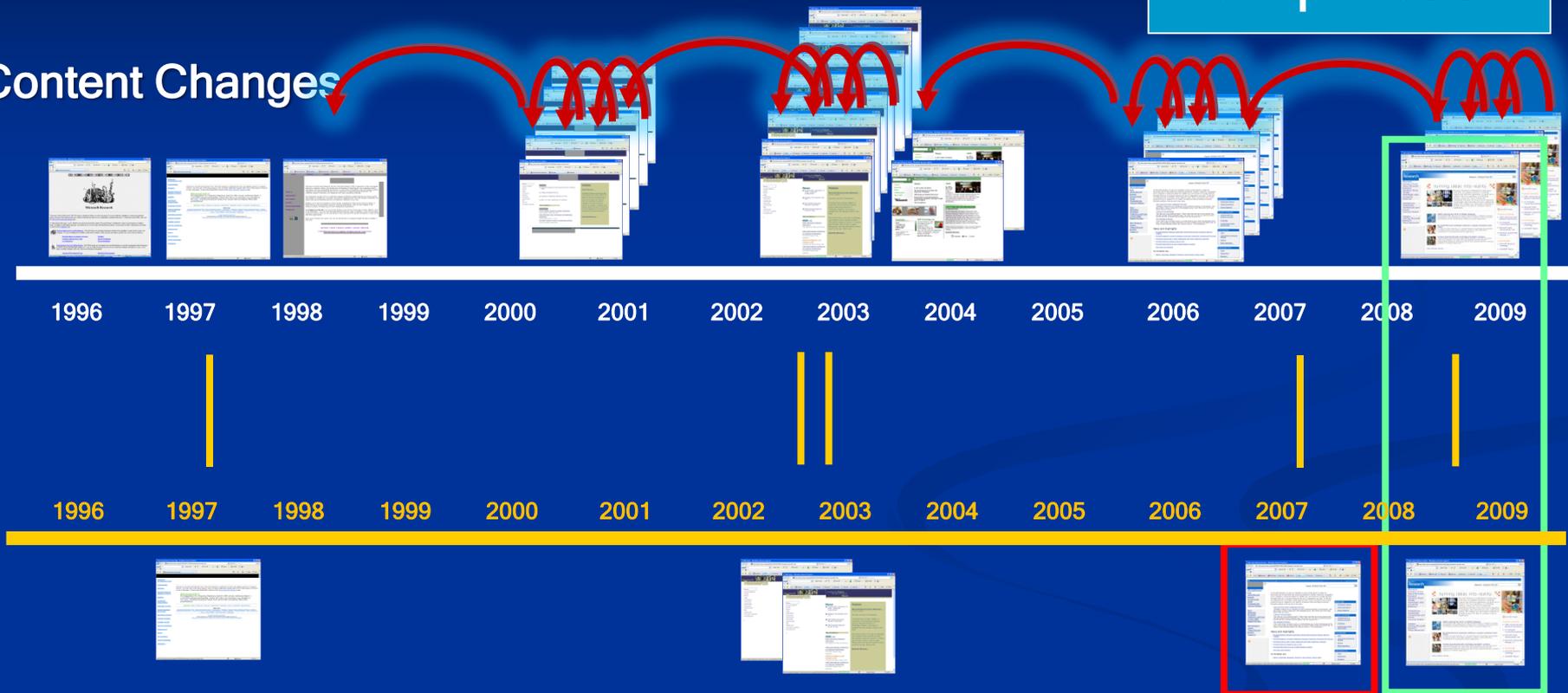
- Explicit annotations, ratings, wikis, etc.
- Implicit interest via interaction patterns
 - Edit wear and read wear (Hill et al., 1992)



Leveraging Dynamics for Retrieval

Temporal IR

Content Changes



User Visitation/ReVisitation

Temporal Retrieval Models

- Current IR 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

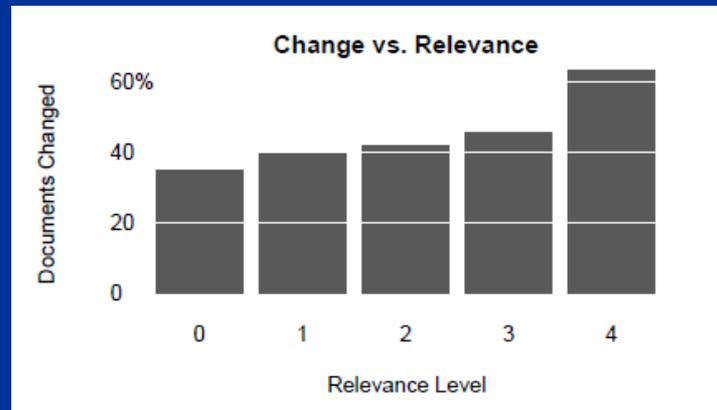
$$P(D|Q) = P(D) \cdot P(Q|D)$$

Change prior

Term longevity

Relevance and Page Change

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



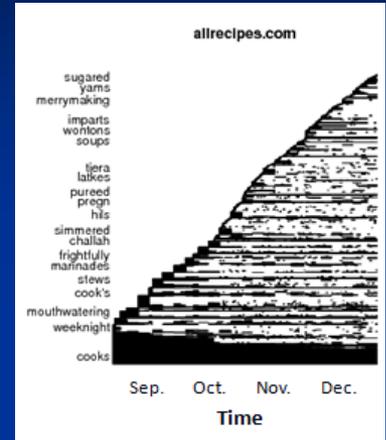
- Use change rate as a document prior (vs. priors based on link structure like Page Rank)
 - Shingle prints to measure change

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

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



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

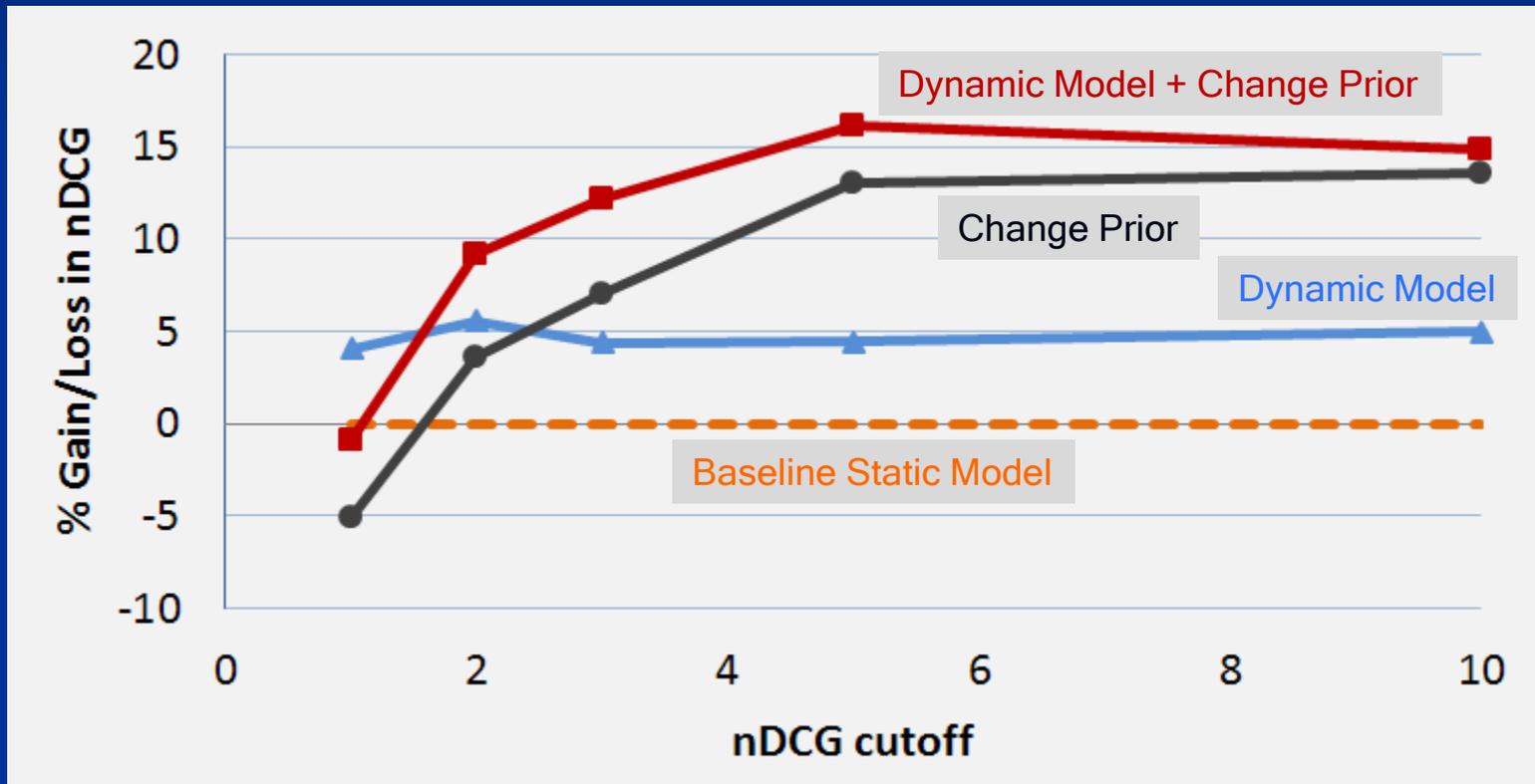
$$P(D | Q) = P(D) \cdot P(Q | D)$$

↑
Term longevity

Evaluation: Queries & Documents

- 18K Queries, 2.5M Judged Documents
 - 5-level relevance judgment (Perfect ... Bad)
- 2.5M Documents crawled weekly for 10 weeks
- Navigational queries
 - 2k queries identified with a “Perfect” judgment
- 60/40 Train/Test split

Experimental Results

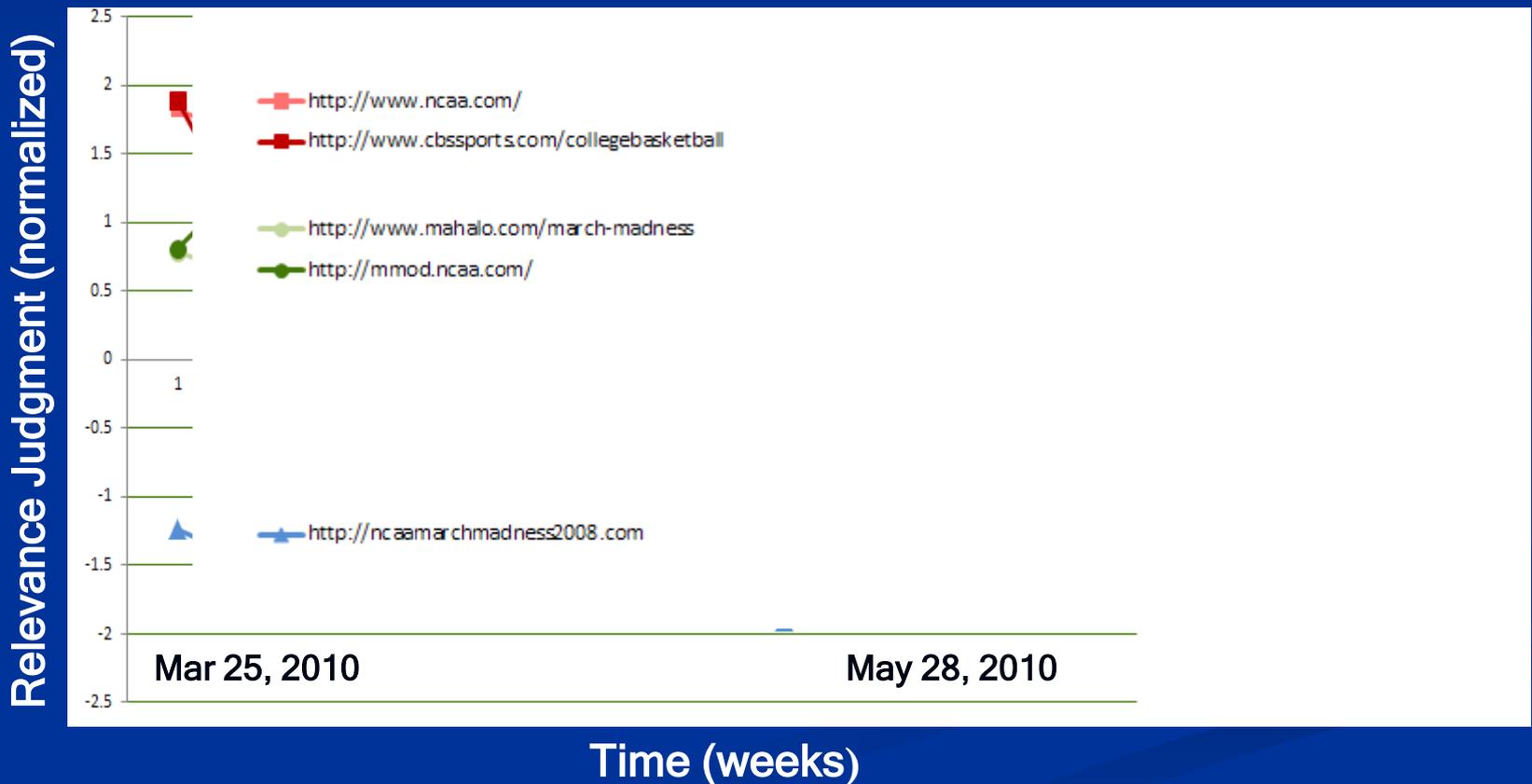


Temporal Retrieval, Ongoing Work

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

Relevance over Time

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



Other Examples of Dynamics and IR Models/Systems

- Temporal retrieval models
 - **Elsas & Dumais (2010)**; Liu & Croft (2004); Efron (2010); Aji et al. (2010)
- Query dynamics
 - Jones & Diaz (2004); Kulkarni et al. (2010); Kotov et al. (2010)
- Integration of news into Web results
 - Diaz (2009)
- Extraction of temporal entities within documents
- Document dynamics, for crawling and indexing
 - Efficient crawling and updates
- Protocols for retrieving versions over time
 - E.g., Memento (Von de Sompel et al., 2010)

Summary

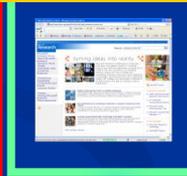
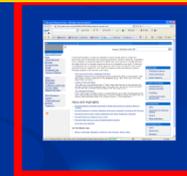
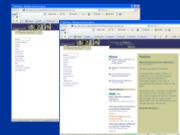
Temporal IR:
Leverages change
for improved IR

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

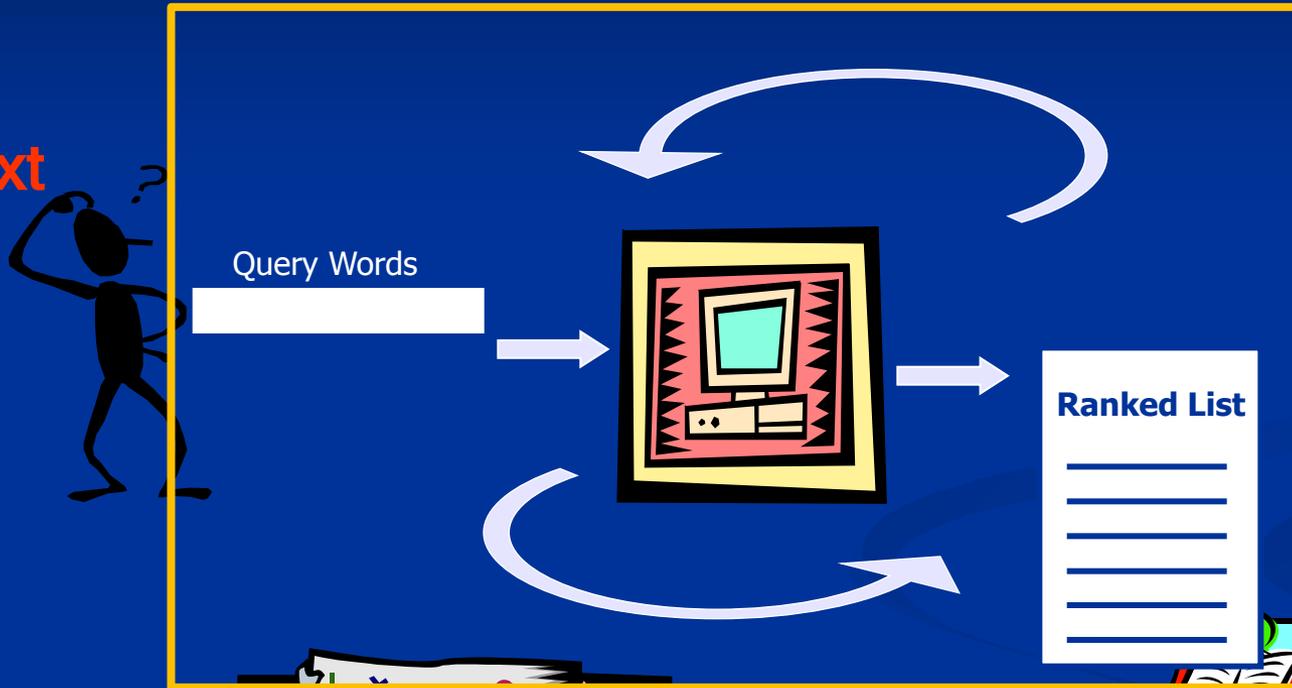
Diff-IE: Supports (and influences) interaction and understanding

Opportunities and Challenges

- Temporal dynamics are pervasive in information systems
- Influence many aspect of IR systems
 - Systems: protocols, indexing versioned collections, crawling
 - Document representations: sufficient statistics at page and term-level, meta-data generation and information extraction
 - Retrieval models: term weights, document priors, etc.
 - Evaluation
 - User experience
 - Analysis and prediction
- Time is one important example of context and IR
 - Others include: location, individuals, tasks ...

Search and Context

User Context



Document Context

Task/Use Context



Thank You !

- Questions/Comments ...
- More info,
<http://research.microsoft.com/~sdumais>

Diff-IE (coming soon),

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



References

Change and Revisitation:

- Adar, Teevan, Dumais & Elsas. The Web changes everything: Understanding the dynamics of Web Content. *WSDM'09*.
- Adar, Teevan & Dumais. Large scale analysis of Web revisitation patterns. *CHI'08*.
- Teevan, Adar, Jones & Potts. Information re-retrieval: Repeat queries in Yahoo's logs. *SIGIR'07*.
- Tyler & Teevan. Large scale query log analysis of re-finding. *WSDM'10*.
- Adar, Teevan & Dumais. Resonance on the Web: Web dynamics and revisitation patterns. *CHI'09*.

Diff-IE and Temporal IR:

- Teevan, Dumais, Liebling & Hughes. Changing how people view changes on the Web. *UIST'09*.
- Teevan, Dumais & Liebling. A longitudinal study of how highlighting Web content change affects people's Web interactions. *CHI'10*.
- Elsas & Dumais. Leveraging temporal dynamics of document content in relevance ranking. *WSDM'10*.
- Kulkarni, Teevan, Svore & Dumais. Understanding temporal dynamic queries. To appear in *WSDM'11*.