Dynamic Information Environments

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

In collaboration with:
Jaime Teevan, Eytan Adar, Jon Elsas, Dan Liebling, Richard Hughes
Outline

- Web search and context
- Temporal dynamics of information
  - Characterizing change
    - Content changes over time
    - People re-visit and re-find
    - Relationships between content change and re-access
  - Improving retrieval and understanding
    - Building support for understanding change (e.g., DiffIE)
    - Leveraging dynamics for improved retrieval
Web Search at 15

What’s available

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

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

How it’s accessed

- Search
  - Search the Web:
  - Google Search
  - Bing

UW, CSE 454, Dec 8 2009
Support for Searchers

- The search box
- Spelling suggestions
- Query suggestions
- Auto complete
- Inline answers
- Richer snippets
- But, we can do better

UW, CSE 454, Dec 8 2009
**Modeling Users**
- Short vs. long term
- Individual vs. group
- Implicit vs. explicit

**Using User Models**
- Stuff I’ve Seen (re-finding)
- Personalized Search
- News Junkie (novelty)
- User Behavior in Ranking
- Domain Expertise at Web-scale

**Systems/Prototypes**
- New capabilities and experiences
- Algorithms and prototypes
- Deploy, evaluate and iterate

**Inter-Relationships among Documents**
- Categorization and Metadata
  - Reuters, spam, landmarks, web categories ...
  - Domain-specific features, time
- Interfaces and Interaction
  - Stuff I’ve Seen, Phlat, Timelines, SWISH
  - Tight coupling of browsing and search

**Search and Context**

**Evaluation**
- Many methods, scales
- Individual components and their combinations

**Temporal Dynamics**

- Redundancy
Information Dynamics

Content Changes

User Visitation/ReVisitation

Today's Browse and Search Experiences

But, ignores ...

UW, CSE 454, Dec 8 2009
Digital Dynamics Easy to Capture

- Easy to capture
- Few tools support dynamics
Information Dynamics

- Characterizing change
  - Content changes over time
  - People re-visit and re-find
  - Relationships between content change and re-access

- Improving retrieval and understanding
  - Building support for understanding change (e.g., DiffIE)
  - Leveraging dynamics for improved retrieval
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
  - Judged pages (Relevance to a query)
    - 6 million pages crawled every two days for 6 months
Measuring Web Page Change

- Summary metrics
  - Number of changes
  - Time between changes
  - Amount of change

- Change curves
  - Fixed starting point
  - Measure similarity over different time intervals
Measuring Within-Page Change

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

[Graph showing time periods and changes]
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

What’s 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
## Search and Revisititation

- **Repeat query (33%)**
  - *microsoft research*

- **Repeat click (39%)**
  - [http://research.microsoft.com](http://research.microsoft.com)
  - Q: *microsoft research, msr* ...

- **Big opportunity (43%)**
  - 24% “navigational revisits”

<table>
<thead>
<tr>
<th></th>
<th>Repeat Click</th>
<th>New Click</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Repeat Query</strong></td>
<td>33%</td>
<td>29%</td>
</tr>
<tr>
<td><strong>New Query</strong></td>
<td>67%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>39%</td>
<td>61%</td>
</tr>
</tbody>
</table>
Repeat Clicks for Repeat Queries

- Within session: Repeat query → New click
- Across sessions: Repeat query → Repeat click
Why did you revisit the last Web page you revisited?
Possible Relationships

- Interested in change
  - Monitor
- Effect change
  - Transact
- Change unimportant
  - Find new
- Change can interfere
  - Re-find
Understanding the Relationship

- **Compare summary metrics**
  - Revisits: Unique visitors, visits/user, interval
  - Change: Number, interval, Dice

<table>
<thead>
<tr>
<th></th>
<th>Number of changes</th>
<th>Time between changes</th>
<th>Dice coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 visits/user</td>
<td>172.91</td>
<td>133.26</td>
<td>0.82</td>
</tr>
<tr>
<td>3 visits/user</td>
<td>200.51</td>
<td>119.24</td>
<td>0.82</td>
</tr>
<tr>
<td>4 visits/user</td>
<td>234.32</td>
<td>109.59</td>
<td>0.81</td>
</tr>
<tr>
<td>5 or 6 visits/user</td>
<td>269.63</td>
<td>94.54</td>
<td>0.82</td>
</tr>
<tr>
<td>7+ visits/user</td>
<td>341.43</td>
<td>81.80</td>
<td>0.81</td>
</tr>
</tbody>
</table>
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 revisited at different rates

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

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, conducting 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.

2008 BEIJING OLYMPICS

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.

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.

THEATER

With Echoes of Chekhov

A.R. Gurney’s “Buffalo Gal” is about an actress who retracts the boards of her hometown.

INTERACTIVE SCHEDULE

MORE OLYMPICS NEWS

- Men’s Soccer: U.S. 1, Japan 0
- Bush Arrives in Beijing
- Silver Keeps Her Place on U.S. Relay Team
- Federer Avoids, Soon to Lose His No. 1 Status
- Rings Blog
- Complete Coverage

Judge Orders Detroit Mayor Jailed

By NICK BUNKLEY 39 minutes ago
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 word? Maybe looking for um, pee or ah are on?
Dynamics of Information

- Characterizing change
  - Content changes over time
  - People re-visit and re-find
  - Relationships between content change and re-access

- Improving retrieval and understanding
  - Building support for understanding change (e.g., DiffIE)
  - Leveraging dynamics for improved retrieval
Interesting Features of **DiffIE**

- **Always on**
- **In-situ**
- **New to you**
- **Non-intrusive**

---

2009. During his career at Microsoft, Gates held the positions of CEO and chief software architect. Gates is one of the best-known entrepreneurs of the personal computer revolution. Although he is admired by many, a number of industry insiders criticize his business tactics, which they consider anti-competitive, an opinion which has in some cases been upheld by the courts (see Criticism of Microsoft). In the later stages of his career, Gates has pursued a number of philanthropic endeavors, donating large amounts of money to various charitable organizations and scientific research programs through the Bill & Melinda Gates Foundation, established in 2000. Bill Gates stepped down as chief executive officer (CEO) of Microsoft in 2000 and took on the role of chairman. The following year, he began serving as chairman of the board of trustees of the Smithsonian Institution, a post he held until 2014.
Examples of DiffIE in Action
Expected New Content
The SIGIR 2009 conference opens in a few days in Boston, Massachusetts, at the Sheraton Boston Hotel and Northeastern University. The conference is chock full of exciting events and registrations are strong and still growing. We are looking forward to an exciting week.

For attendees
- Visas
- Boston
- Venue
- Registration
  - student support
  - accommodations
  - Red Sox

Recent news and upcoming deadlines
- Please join your colleagues by starting the conference with a free continental breakfast in the Sheraton Hotel, Back Bay A&B, from 7:00am to 8:20am on Monday, July 20.
- The conference banquet is currently full. Effective July 16th, reservations will be wait-listed for the banquet.
- Standard conference registration closes the night of Sunday, July 19th (Boston time). See below for more information.
- Registration for Northeastern University housing has closed. If you need to change your reservation, contact questions@sigir2009.org.
- Several special tourism events have been added to the schedule: a panoramic view of the city on Sunday, a famous duck boat tour on Saturday, and a sunset harbor cruise on Wednesday. Look under “Boston” on the left for more information. Two of the events require sufficient interest to occur, so please fill out the questionnaires on that page.
- Industry Track registration is now available for students.

Arrival at the conference
- Tutorial and doctoral consortium registration is on the Northeastern campus, Sunday morning (9:30). If you have registered for a tutorial or want to register for a tutorial, please get to Smith Hall between 8:30 and 9:00 (the doctoral consortium starts at 8:30, morning tutorials start at 9:00). Volunteers and signs will provide direction from the hotel and the NEU dorms. Here is a map.
- Normal conference registration takes place at the Sheraton hotel, starting at 3:00pm on Sunday, July 19. Registration will be available throughout the conference if you are not arriving on Sunday.
- Workshop registrations may be made during normal registration periods or on the Northeastern campus on Thursday morning (the 23rd).
- If you have paid your fees in full in advance, your registration process should be very speedy.
- A free breakfast is available Monday morning (only) in the Sheraton.

For those who are driving, parking is available at the Sheraton and at Northeastern University. See here
Serendipitous Encounters

Organized by Type * Subject * Date * Internal

RECOMMENDED JOURNAL ARTICLES:


RECOMMENDED CONFERENCE PAPERS:


### Most Downloaded

<table>
<thead>
<tr>
<th>App Name</th>
<th>Description</th>
<th>Downloads</th>
<th>Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>SnowSenderPhoto</td>
<td>Automatically displays a photo of every Microsoft Employee that sends you mail.</td>
<td>1,676</td>
<td>01/15/2008</td>
</tr>
<tr>
<td>Bugger</td>
<td>Bugger is a personal bug portal, designed to make your bug list easier to manage and a bit more fun.</td>
<td>1,093</td>
<td>06/13/2006</td>
</tr>
<tr>
<td>Magic Mouse</td>
<td>A &quot;virtual&quot; keyboard/mouse to control multiple machines concurrently, copy/paste drag/drop test/media/files across machines.</td>
<td>960</td>
<td>06/12/2008</td>
</tr>
<tr>
<td>Outlook Conference Room Finder</td>
<td>Find an available conference room in a specific floor and building at any time from within the new appointment dialog in Outlook!</td>
<td>833</td>
<td>03/13/2005</td>
</tr>
<tr>
<td>Source Depot Browser (SDB)</td>
<td>GUI client for Source Depot</td>
<td>613</td>
<td>02/08/2005</td>
</tr>
<tr>
<td>RDMan</td>
<td>RDMan manages multiple remote desktop connections.</td>
<td>577</td>
<td>02/08/2005</td>
</tr>
<tr>
<td>RichCopy 4.0.211</td>
<td>RichCopy is designed to help people who copy huge number of files between storage devices.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### New Releases

<table>
<thead>
<tr>
<th>App Name</th>
<th>Description</th>
<th>Released</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploitable</td>
<td>A Windows Debugging Engine extension for auto-tracing of exceptions based on their security imp... more</td>
<td>04/15/2009</td>
</tr>
<tr>
<td>Biztalk MsgBox Viewer</td>
<td>Health Check tool for Biztalk Databases</td>
<td>04/15/2009</td>
</tr>
<tr>
<td>Send-a-Bug</td>
<td>Send-a-Bug allows you to easily send Product Studio bugs in Outlook. It automatically adds the... more</td>
<td>04/15/2009</td>
</tr>
<tr>
<td>Fix Build Order</td>
<td>FixBuildOrder ensures the build order is correct for a SLN file that contains csproj or vbproj... more</td>
<td>04/15/2009</td>
</tr>
<tr>
<td>Machine Management Tool (MMT)</td>
<td>The Machine Management Tool is a simple C# application written to manage machines in the lab e... more</td>
<td>04/14/2009</td>
</tr>
</tbody>
</table>

### Top Rated

<table>
<thead>
<tr>
<th>App Name</th>
<th>Description</th>
<th>Rating</th>
<th>Released</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clip</td>
<td>Command line Windows clipboard operations</td>
<td>5.0</td>
<td>02/08/2005</td>
</tr>
<tr>
<td>ERR</td>
<td>command-line-based error lookup tool</td>
<td>5.0</td>
<td>02/08/2005</td>
</tr>
<tr>
<td>Shist</td>
<td>Source history browser and function tracker tool</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Understand Page Dynamics

10 Day Weather Forecast for Bellevue, WA (98005) - weather.com
10 day weather forecast for Bellevue, WA (98005) with weather conditions, high and low temperatures, and precipitation.
www.weather.com/weather/tenday/98005?day=0&vert=WeatherCity · Cached page

Bellevue, WA (98006) Weather Forecast - Weatherlet.com
Bellevue, WA (98006) Weather Forecast - Weatherlet ... Auburn, WA (98002)
www.weatherlet.com/?locid=98006&cityname=Bellevue%2C+today... · Cached page

Bellevue, Washington (98005) Conditions & Forecast : Weather ...
Weather Underground provides weather information for worldwide locations, including current conditions, hourly forecasts, radar and satellite maps. Specialized weather products ...
## Attend to Activity

- '13: Eastside Mothers & More Expo. See Chapter News and Events for more information. Yeah, baby!!

<table>
<thead>
<tr>
<th>Message Board</th>
<th>Discussion Forums</th>
<th>Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1] 2 3 ... 9</strong></td>
<td>Mark Topics as Read for this Board</td>
<td>Notify of replies</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td><strong>Started by</strong></td>
<td><strong>Replies</strong></td>
</tr>
<tr>
<td>Small Business « 1 2 »</td>
<td>Della Dawn</td>
<td>21</td>
</tr>
<tr>
<td>Defining Sequencing</td>
<td>Miti Morris</td>
<td>0</td>
</tr>
<tr>
<td>Back to work Part Time! new</td>
<td>Sue Monahan</td>
<td>8</td>
</tr>
<tr>
<td>Aldersgate Christian Preschool - HIRING new</td>
<td>Heather Ross</td>
<td>0</td>
</tr>
<tr>
<td>Do you know anyone in Italy or Sweden?</td>
<td>Virginia Menting</td>
<td>3</td>
</tr>
<tr>
<td>Got an interview and I'm freaking out... « 1 2 ... 6 7 »</td>
<td>Trina Sooy</td>
<td>94</td>
</tr>
<tr>
<td>Academia or Industry? Finding the Right Fit</td>
<td>Jaime Teeevar</td>
<td>7</td>
</tr>
<tr>
<td>Any members work at Amazon.com?</td>
<td>Kathryn Pizzo</td>
<td>10</td>
</tr>
<tr>
<td>Time Magazine on SAHMs getting back into workforce</td>
<td>Kathryn Pizzo</td>
<td>2</td>
</tr>
<tr>
<td>Freelance Resource</td>
<td>Kathryn Pizzo</td>
<td>0</td>
</tr>
<tr>
<td>Graphic designer or Admin?</td>
<td>Heather Ross</td>
<td>0</td>
</tr>
</tbody>
</table>
Edit

Installing DiffIE

- Installation location: \asibuild\Drop\CLUES\latest\DiffIE\DiffIE.msi
  Note: Save this MSI to your machine and stop IE before installing DiffIE.
- Announcement alias: If you download DiffIE, please join the “DiffIE Announcements” distribution list so that we can keep you informed about new versions and other important changes.
- Feedback alias: diffieib@microsoft.com

Using DiffIE

Once installed, you should see the DiffIE toolbar in IE (shown below on Vista):
Expected

- New Content
- Monitor

Unexpected

- Important Content
- Serendipitous Encounter
- Unimportant Content

Unimportant Content

Attend to Activity

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT

Monitor

Attend to Activity

Serendipitous Encounter

Unimportant Content

Edit

Understand

Page Dynamics

Expected

Unusual

NEW CONTENT
Methods for Studying DiffIE

- Feedback buttons
- Survey
  - Prior to installation
  - After a month of use
- 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
  - First week: 39.4% of visits are revisits
  - Last week: 45.0% of visits are revisits

- Why are people revisiting more with DiffIE?
Revisited Pages Change More

- Perception of change increases
  - What proportion of pages change regularly?
  - How often do you notice unexpected change?

- Amount of change seen increases
  - First week: 21.5% revisits changed by 6.2%
  - Last week: 32.4% revisits changed by 9.5%

- *DiffIE* is driving visits to changed pages
Change by Page Type

- Perceptions of change reinforced
- Pages that change a lot $\rightarrow$ change more
- Pages that change a little $\rightarrow$ change less

- News pages
- Message boards, forums, news groups
- Search engine results
- Blogs you read
- Pages with product information
- Wikipedia pages
- Company homepages
- Personal home pages of people you know
- Reference pages (dictionaries, yellow pages, maps)
Leveraging Dynamics for Retrieval

Content Changes


UW, CSE 454, Dec 8 2009
Temporal Retrieval Models

- Current IR algorithms look only at a single snapshot of a page
- But, Web pages change over time
- Can we leverage this for 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
Temporal Retrieval Models

- 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 links, Page Rank)

UW, CSE 454, Dec 8 2009
Temporal Retrieval Models

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

\[ P(Q \mid D) = \lambda_L P(Q \mid D_L) + \lambda_M P(Q \mid D_M) + \lambda_S P(Q \mid D_S) \]

- Improves retrieval accuracy

<table>
<thead>
<tr>
<th>Term Dynamics</th>
<th>Single Slice</th>
<th>Dynamic</th>
<th>nDCG@5 (2k Navigation Queries)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD + Change Prior</td>
<td>0.24</td>
<td>0.25</td>
<td>13.1%</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>0.22</td>
<td>4.4%</td>
</tr>
<tr>
<td></td>
<td>11.2%</td>
<td>19.0%</td>
<td></td>
</tr>
</tbody>
</table>
Summary

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

Temporal IR: Leverage change for improved IR

DiffIE: Support influences interaction and understanding
Think Outside the (Search) Box

User Context
Query Words
Ranked List
Document Context
Task/Use Context
Thank You!

- Questions/Comments ...

- More info, 
  http://research.microsoft.com/~sdumais
References

Change and Revisitation:

- 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.

DiffIE and Temporal IR: