# Leveraging Temporal Dynamics of Document Content in Relevance Ranking

Jonathan L. Elsas (CMU)

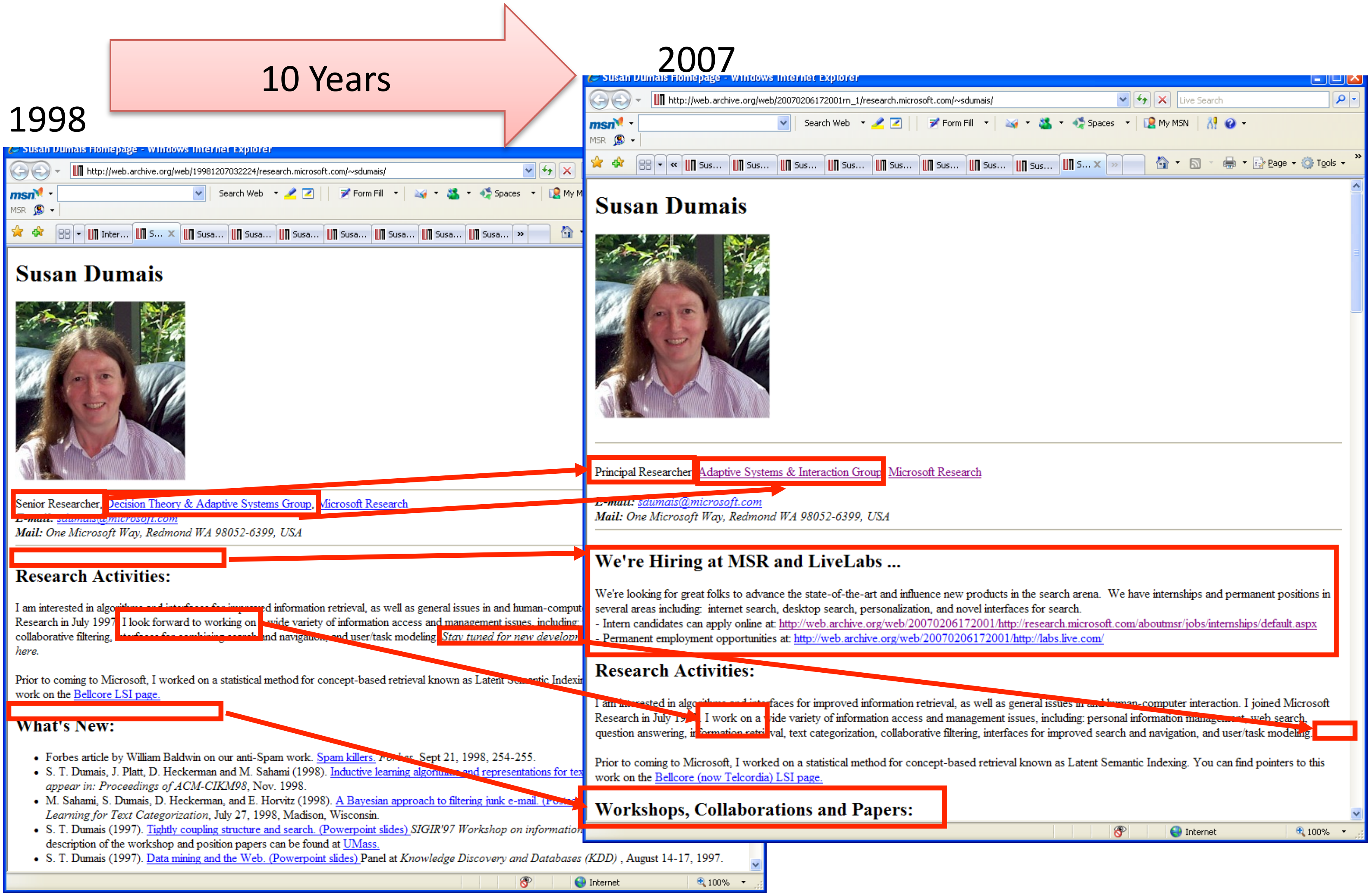
Susan T. Dumais (MSR)

# Outline

- Document Dynamics on the Web
  - Previous Work on Change & What's Missing
- Our Setting: Ranking Dynamic Documents
  - Test Collection & Measuring change
- Two ways to leverage change in ranking
  - Document Prior based on Gross Change Measures
  - Document Representation based on Term-Level Change
- Discussion

# The Web is Dynamic

# The Web is Dynamic


10 Years
1998
2007
Susan Dumais Homepage - Windows Internet Explorer
http://web.archive.org/web/19981207032224/research.microsoft.com/~sdumais/
http://web.archive.org/web/20070206172001rn_1/research.microsoft.com/~sdumais/
Live Search
Search Web
Form Fill
Spaces
My MSN
Page
Tools
Internet
100%

**1998**

## Susan Dumais

Senior Researcher, Decision Theory & Adaptive Systems Group, Microsoft Research
*E-mail:* saumais@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-comput
Research in July 1997. I look forward to working on a wide variety of information access and management issues, including:
collaborative filtering, interfaces for combining search and navigation, and user/task modeling. *Stay tuned for new developn*
*here.*

Prior to coming to Microsoft, I worked on a statistical method for concept-based retrieval known as Latent Semantic Indexin
work on the Bellcore LSI page.

### What's New:

- Forbes article by William Baldwin on our anti-Spam work. Spam killers. *Forbes*, Sept 21, 1998, 254-255.
- S. T. Dumais, J. Platt, D. Heckerman and M. Sahami (1998). Inductive learning algorithms and representations for tex
  *appear in: Proceedings of ACM-CIKM98*, Nov. 1998.
- M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz (1998). A Bayesian approach to filtering junk e-mail. (Post
  *Learning for Text Categorization*, July 27, 1998, Madison, Wisconsin.
- S. T. Dumais (1997). Tightly coupling structure and search. (Powerpoint slides) *SIGIR'97 Workshop on information*
  description of the workshop and position papers can be found at UMass.
- S. T. Dumais (1997). Data mining and the Web. (Powerpoint slides) Panel at *Knowledge Discovery and Databases (KDD)* , August 14-17, 1997.

**2007**

## Susan Dumais

Principal Researcher, Adaptive Systems & Interaction Group, Microsoft Research

*E-mail:* saumais@microsoft.com
*Mail: One Microsoft Way, Redmond WA 98052-6399, USA*

### We're Hiring at MSR and LiveLabs ...

We're looking for great folks to advance the state-of-the-art and influence new products in the search arena. We have internships and permanent positions in several areas including: internet search, desktop search, personalization, and novel interfaces for search.
- Intern candidates can apply online at: http://web.archive.org/web/20070206172001/http://research.microsoft.com/aboutmsr/jobs/internships/default.aspx
- Permanent employment opportunities at: http://web.archive.org/web/20070206172001/http://labs.live.com/

### Research Activities:

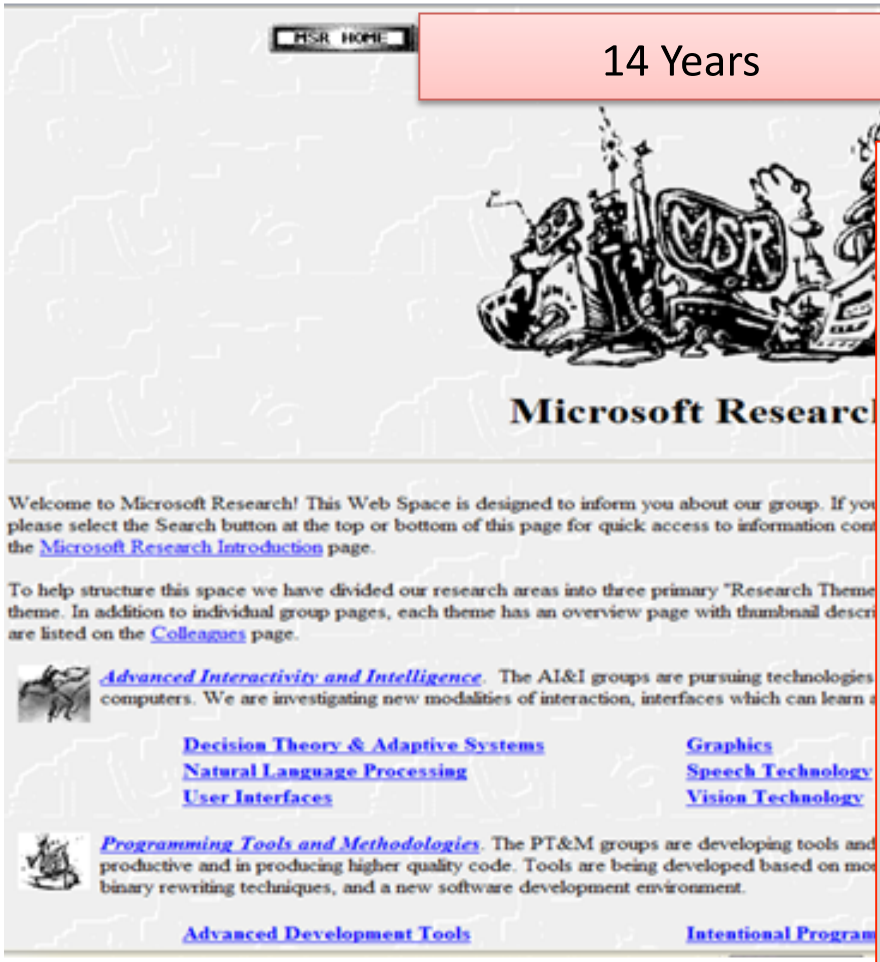
I am interested in algorithms and interfaces for improved information retrieval, as well as general issues in and human-computer interaction. I joined Microsoft Research in July 1997. I work on a wide variety of information access and management issues, including: personal information management, web search, question answering, information retrieval, text categorization, collaborative filtering, interfaces for improved search and navigation, and user/task modeling.

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 (now Telcordia) LSI page.

### Workshops, Collaborations and Papers:

# The Web is Dynamic

1996



MSR HOME

Microsoft Researc

Welcome to Microsoft Research! This Web Space is designed to inform you about our group. If you please select the Search button at the top or bottom of this page for quick access to information con the Microsoft Research Introduction page.

To help structure this space we have divided our research areas into three primary "Research Theme theme. In addition to individual group pages, each theme has an overview page with thumbnail descri are listed on the Colleagues page.

*Advanced Interactivity and Intelligence*. The AI&I groups are pursuing technologies computers. We are investigating new modalities of interaction, interfaces which can learn

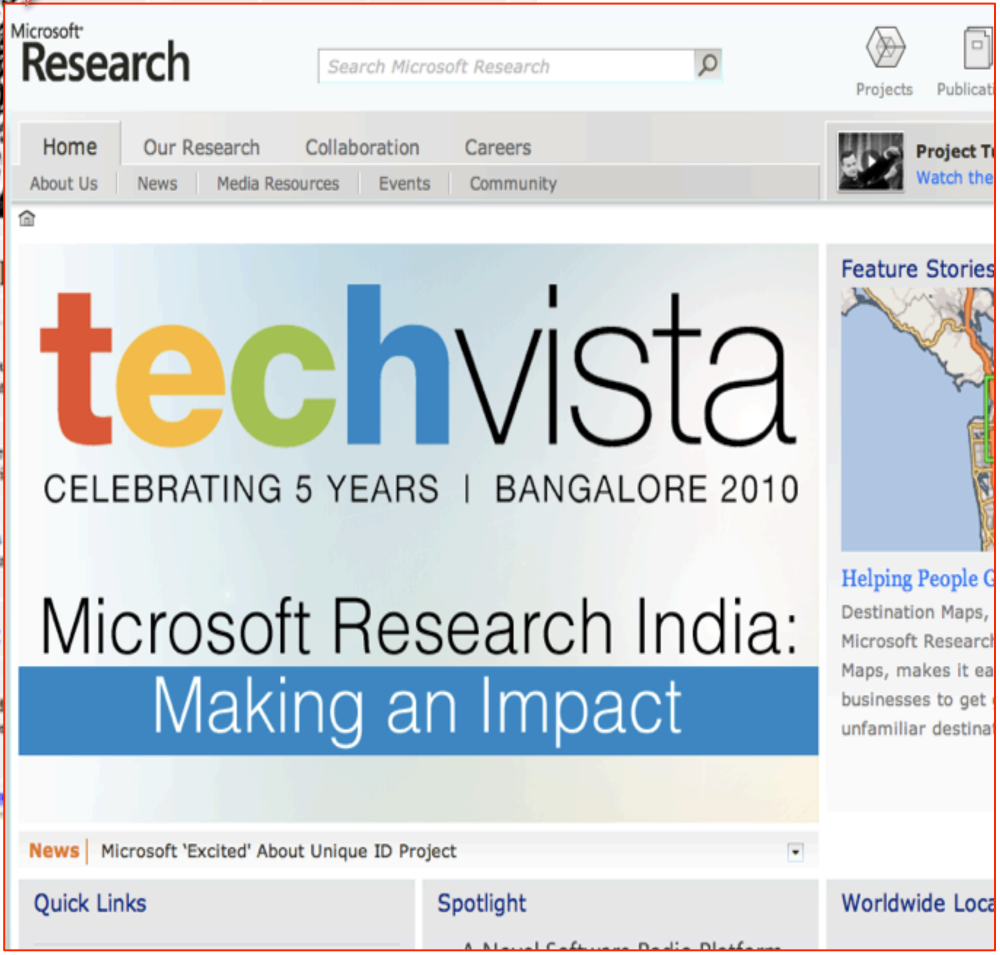
**Decision Theory & Adaptive Systems** | **Graphics**
**Natural Language Processing** | **Speech Technology**
**User Interfaces** | **Vision Technology**

*Programming Tools and Methodologies*. The PT&M groups are developing tools and productive and in producing higher quality code. Tools are being developed based on mo binary rewriting techniques, and a new software development environment.

**Advanced Development Tools** | **Intentional Program**

14 Years

2010



Microsoft Research

Search Microsoft Research

Projects | Publicat

Home | Our Research | Collaboration | Careers

About Us | News | Media Resources | Events | Community

Project T
Watch the

techvista

CELEBRATING 5 YEARS | BANGALORE 2010

Microsoft Research India:
Making an Impact

News | Microsoft 'Excited' About Unique ID Project

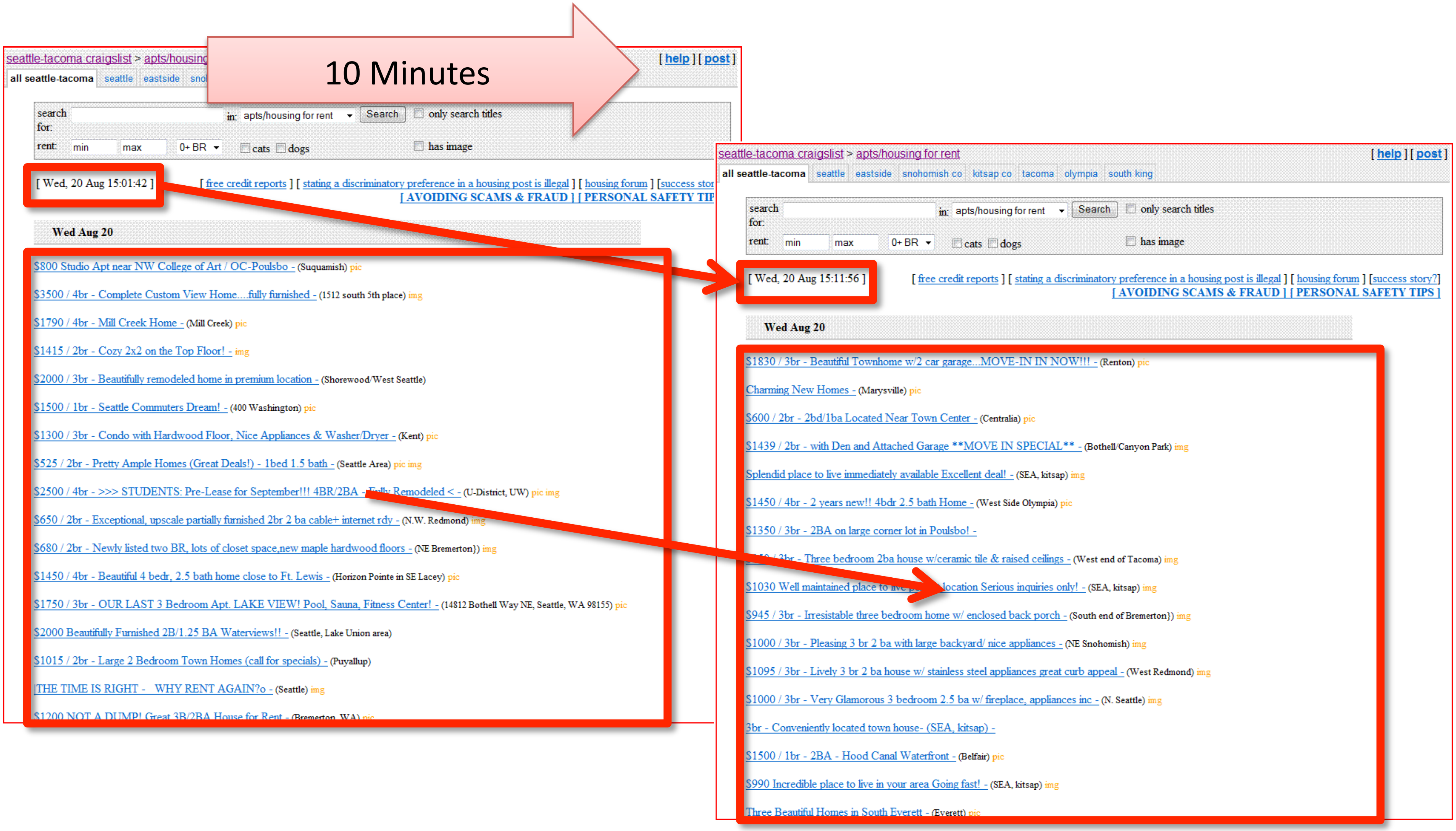
Feature Stories

Helping People G

Destination Maps, Microsoft Researc Maps, makes it ea businesses to get unfamiliar destina

Quick Links | Spotlight | Worldwide Loca

# The Web is Dynamic



seattle-tacoma craigslist > apts/housing

[ help ] [ post ]

all seattle-tacoma | seattle | eastside | sno

search for: in: apts/housing for rent Search □ only search titles

rent: min max 0+ BR □ cats □ dogs □ has image

[ Wed, 20 Aug 15:01:42 ]

[ free credit reports ] [ stating a discriminatory preference in a housing post is illegal ] [ housing forum ] [success stor

[ AVOIDING SCAMS & FRAUD ] [ PERSONAL SAFETY TIP

**Wed Aug 20**

$800 Studio Apt near NW College of Art / OC-Poulsbo - (Suquamish) pic

$3500 / 4br - Complete Custom View Home....fully furnished - (1512 south 5th place) img

$1790 / 4br - Mill Creek Home - (Mill Creek) pic

$1415 / 2br - Cozy 2x2 on the Top Floor! - img

$2000 / 3br - Beautifully remodeled home in premium location - (Shorewood/West Seattle)

$1500 / 1br - Seattle Commuters Dream! - (400 Washington) pic

$1300 / 3br - Condo with Hardwood Floor, Nice Appliances & Washer/Dryer - (Kent) pic

$525 / 2br - Pretty Ample Homes (Great Deals!) - 1bed 1.5 bath - (Seattle Area) pic img

$2500 / 4br - >>> STUDENTS: Pre-Lease for September!!! 4BR/2BA - Fully Remodeled < - (U-District, UW) pic img

$650 / 2br - Exceptional, upscale partially furnished 2br 2 ba cable+ internet rdy - (N.W. Redmond) img

$680 / 2br - Newly listed two BR, lots of closet space,new maple hardwood floors - (NE Bremerton}) img

$1450 / 4br - Beautiful 4 bedr, 2.5 bath home close to Ft. Lewis - (Horizon Pointe in SE Lacey) pic

$1750 / 3br - OUR LAST 3 Bedroom Apt. LAKE VIEW! Pool, Sauna, Fitness Center! - (14812 Bothell Way NE, Seattle, WA 98155) pic

$2000 Beautifully Furnished 2B/1.25 BA Waterviews!! - (Seattle, Lake Union area)

$1015 / 2br - Large 2 Bedroom Town Homes (call for specials) - (Puyallup)

|THE TIME IS RIGHT - WHY RENT AGAIN?o - (Seattle) img

$1200 NOT A DUMP! Great 3B/2BA House for Rent - (Bremerton, WA) pic

seattle-tacoma craigslist > apts/housing for rent

[ help ] [ post ]

all seattle-tacoma | seattle | eastside | snohomish co | kitsap co | tacoma | olympia | south king

search for: in: apts/housing for rent Search □ only search titles

rent: min max 0+ BR □ cats □ dogs □ has image

[ Wed, 20 Aug 15:11:56 ]

[ free credit reports ] [ stating a discriminatory preference in a housing post is illegal ] [ housing forum ] [success story?]

[ AVOIDING SCAMS & FRAUD ] [ PERSONAL SAFETY TIPS ]

**Wed Aug 20**

$1830 / 3br - Beautiful Townhome w/2 car garage...MOVE-IN IN NOW!!! - (Renton) pic

Charming New Homes - (Marysville) pic

$600 / 2br - 2bd/1ba Located Near Town Center - (Centralia) pic

$1439 / 2br - with Den and Attached Garage **MOVE IN SPECIAL** - (Bothell/Canyon Park) img

Splendid place to live immediately available Excellent deal! - (SEA, kitsap) img

$1450 / 4br - 2 years new!! 4bdr 2.5 bath Home - (West Side Olympia) pic

$1350 / 3br - 2BA on large corner lot in Poulsbo! -

[illegible] / 3br - Three bedroom 2ba house w/ceramic tile & raised ceilings - (West end of Tacoma) img

$1030 Well maintained place to live p [illegible] location Serious inquiries only! - (SEA, kitsap) img

$945 / 3br - Irresistable three bedroom home w/ enclosed back porch - (South end of Bremerton}) img

$1000 / 3br - Pleasing 3 br 2 ba with large backyard/ nice appliances - (NE Snohomish) img

$1095 / 3br - Lively 3 br 2 ba house w/ stainless steel appliances great curb appeal - (West Redmond) img

$1000 / 3br - Very Glamorous 3 bedroom 2.5 ba w/ fireplace, appliances inc - (N. Seattle) img

3br - Conveniently located town house- (SEA, kitsap) -

$1500 / 1br - 2BA - Hood Canal Waterfront - (Belfair) pic

$990 Incredible place to live in your area Going fast! - (SEA, kitsap) img

Three Beautiful Homes in South Everett - (Everett) pic

# Previous Work on Dynamics

Characterizing Change

Implications of Change

# Characterizing Change

Change & Page Type / Source

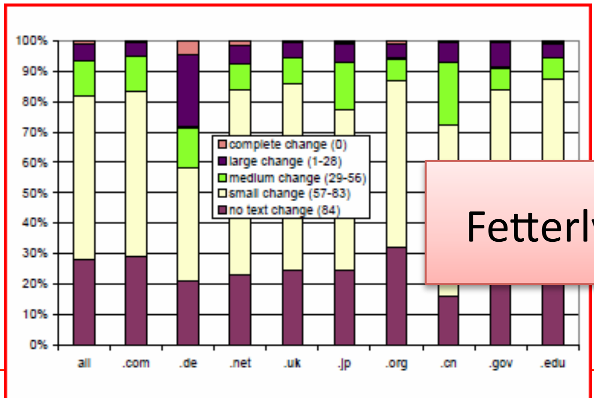
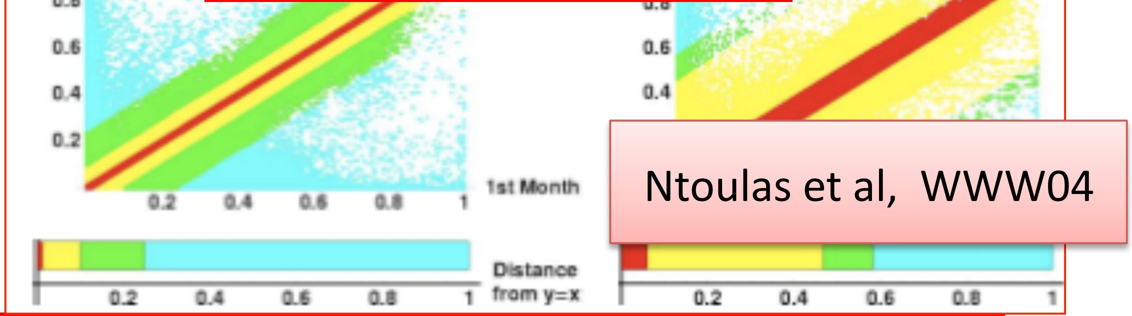


**Figure 12:** Clustered rates of change, broken down by selected top-level domains, and omitting the *no change* cluster.

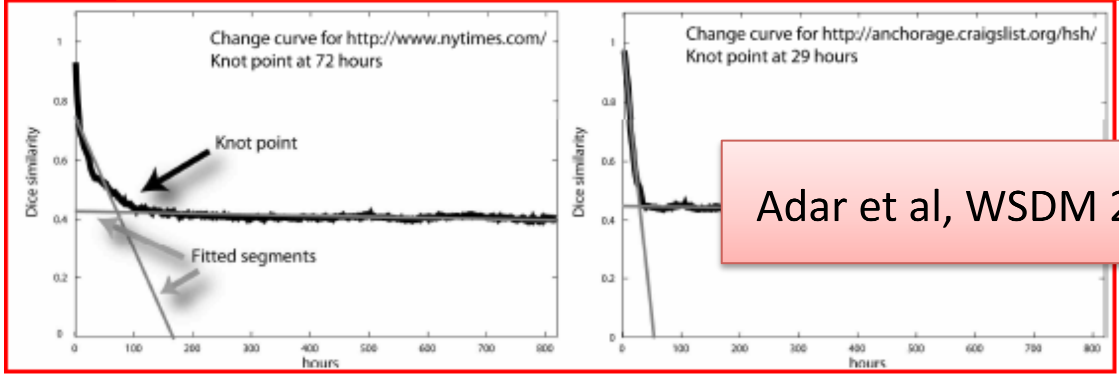
Fetterly et al, WWW03

New Content & Links



Ntoulas et al, WWW04

Within-Doc. Change



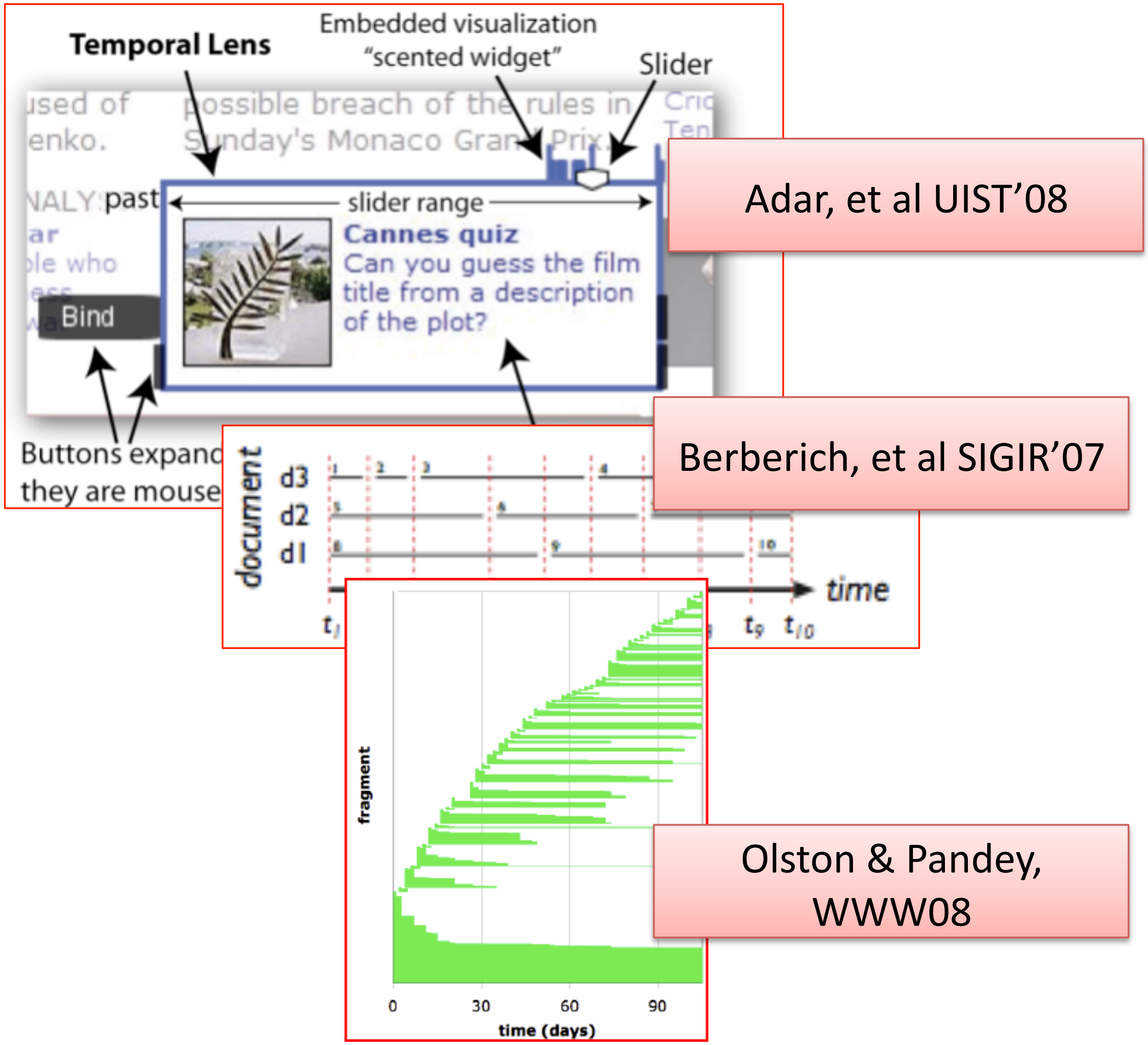
Adar et al, WSDM 2009

# Implications of Change

On Browsing

On Indexing

On Crawling



Adar, et al UIST'08

Berberich, et al SIGIR'07

Olston & Pandey, WWW08

# Implications of Change

On Ranking?

# Implications of Change on Ranking

- Gross Measures of Document Change

  Are there general characteristics of document dynamics indicate high quality pages?

- Representing Term-Level Change Within the Document

  Are there characteristics of a document's dynamic content that indicate some content may be more important?

# Test Setup: Queries & Documents

- 18K Queries, 2.5M Judged Documents
  - 5-level relevance judgment (Bad…Perfect)
- 2.5M Documents crawled weekly for 10 weeks

- Navigational queries
  - 2k queries identified with a “Perfect” judgment
- 60/40 Training/Test split

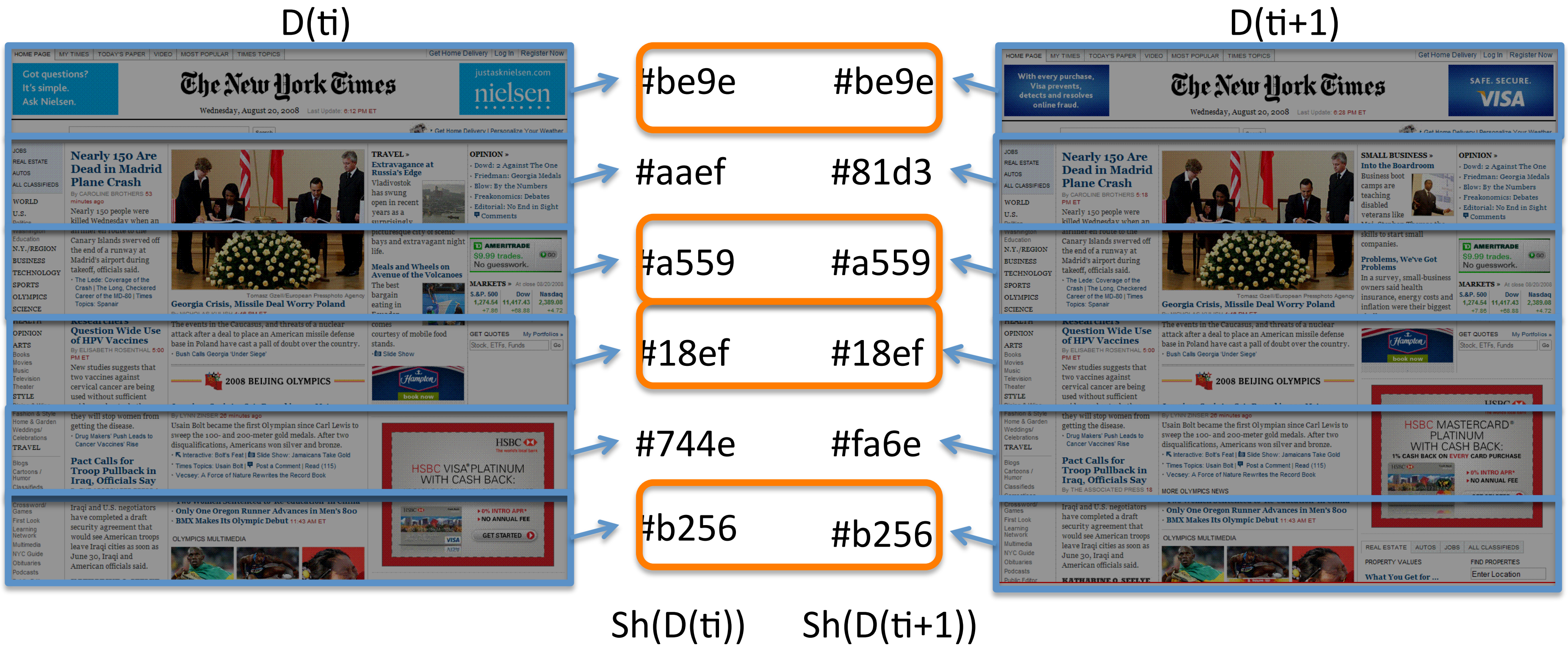
# Test Setup: Queries & Documents

- 18K Queries, 2.5M Judged Documents
  - 5-level relevance
- 2.5M Documents

We focus on Navigational Queries here for ease of evaluation.

- Navigational queries
  - 2k queries identified with a “Perfect” judgment
- 60/40 Training/Test split

# Measuring Change: Shingleprints



D(ti)

D(ti+1)

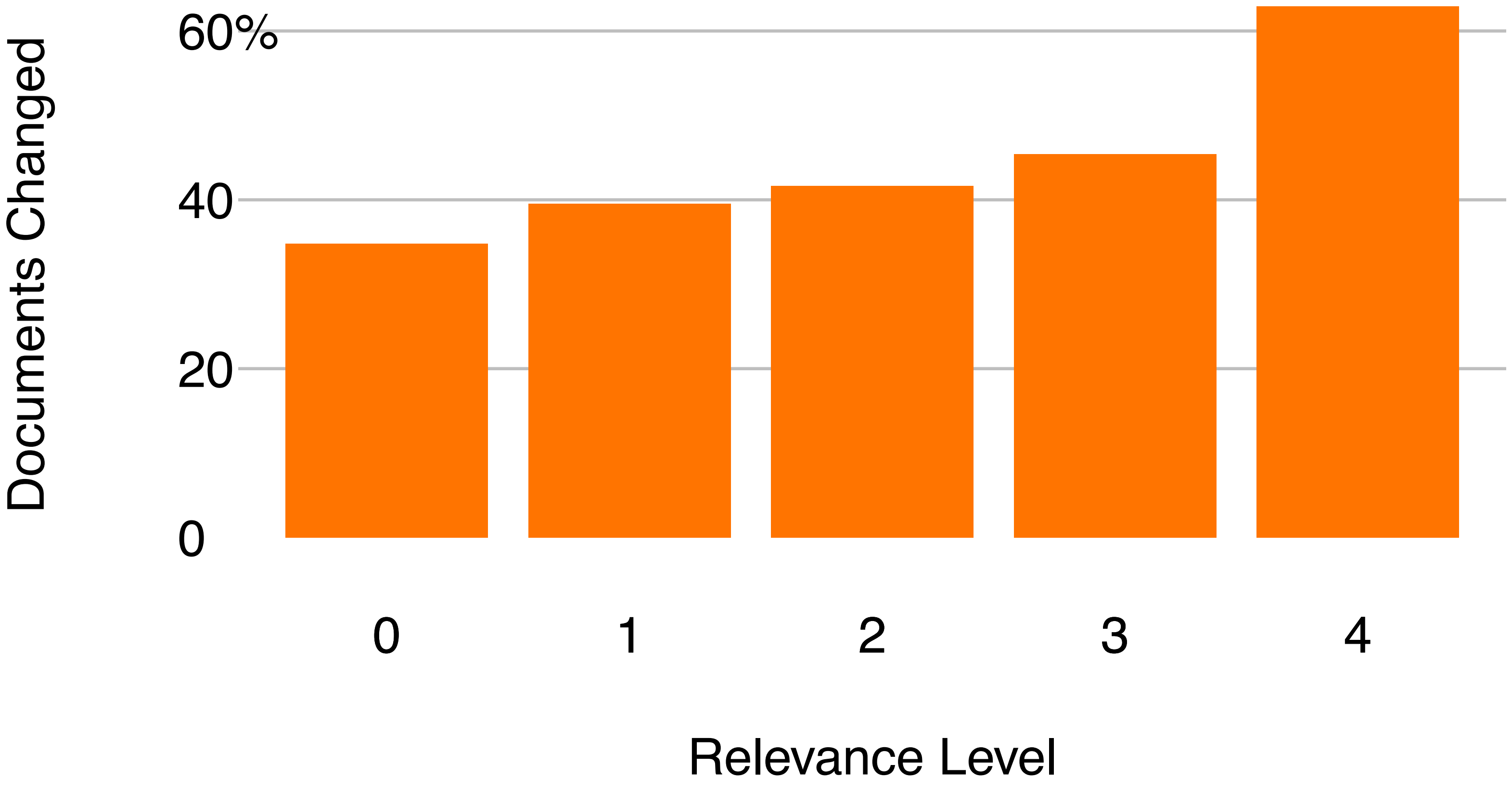
| | |
|---|---|
| #be9e | #be9e |
| #aaef | #81d3 |
| #a559 | #a559 |
| #18ef | #18ef |
| #744e | #fa6e |
| #b256 | #b256 |

Sh(D(ti)) Sh(D(ti+1))

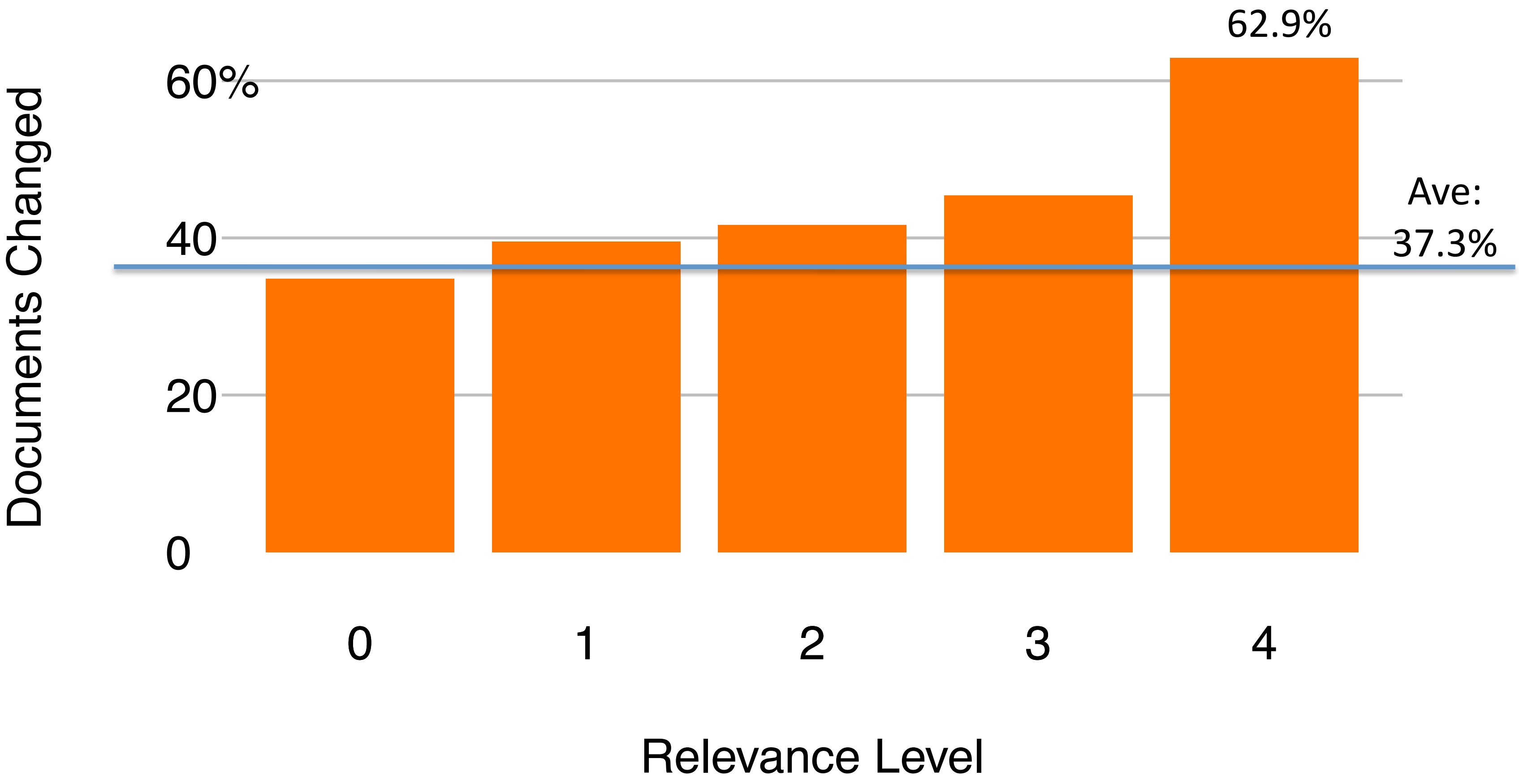
$$ShDiff(D) = 1 - \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{|Sh(D^{(t)}) \cap Sh(D^{(t+1)})|}{N}$$

Broder, et al, “Syntactic Clustering of the Web”
Computer Networks & ISDN Sys., 1997

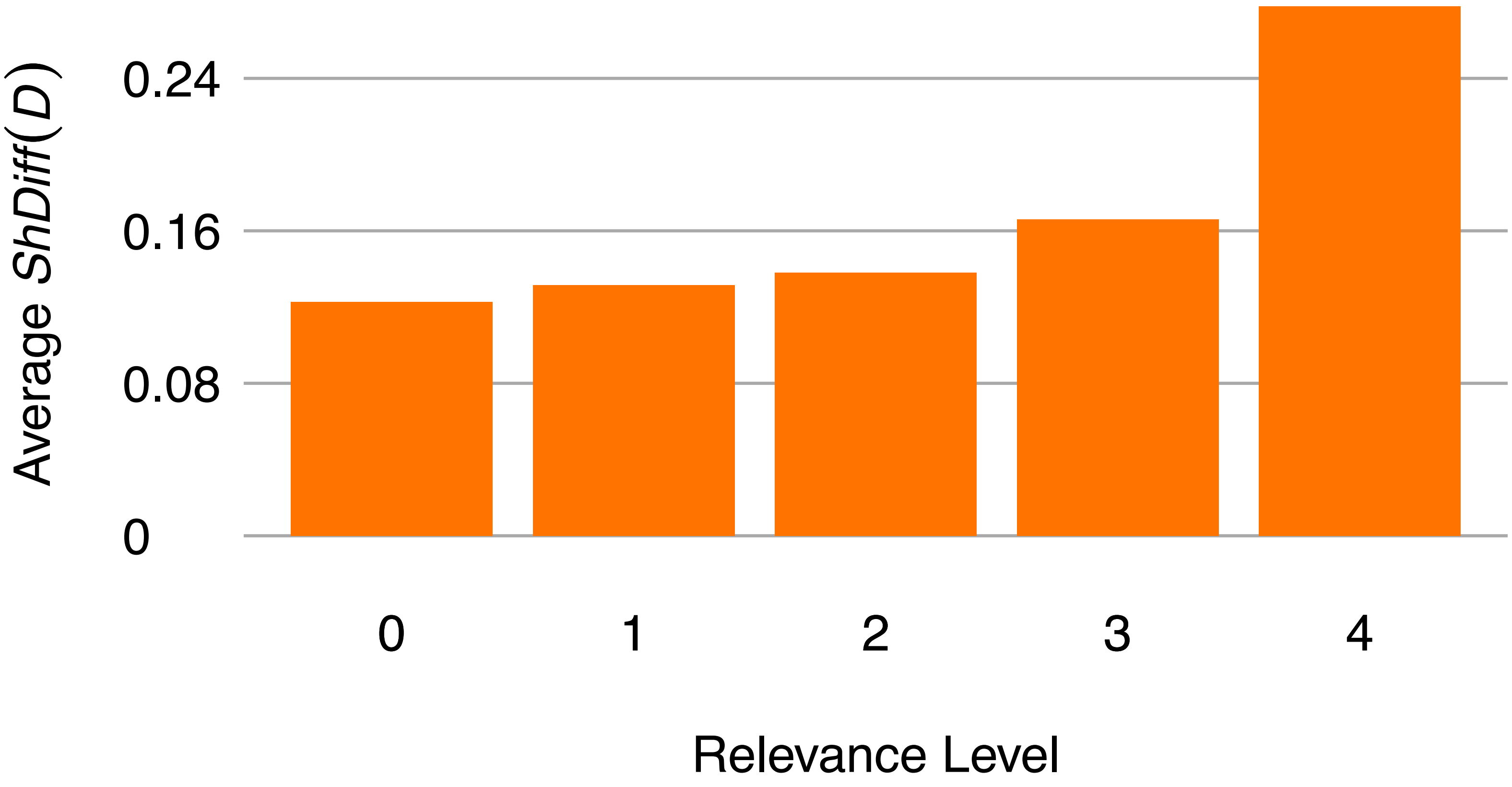
# Change & *Relevance*



# Change & *Relevance*



# Change Amount & *Relevance*



# Change & *Relevance*

- More relevance documents tend to change more often, *and* to a greater degree than non-relevant documents.

- Could favoring dynamic documents in ranking improve performance?

# Favoring Dynamic Documents

Language-Modeling Ranking Function:

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

# Favoring Dynamic Documents

Language-Modeling Ranking Function:

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

Uniform Prior:

$$P(D) \propto 1.0$$

“Static Model”

# Favoring Dynamic Documents

Language-Modeling Ranking Function:

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

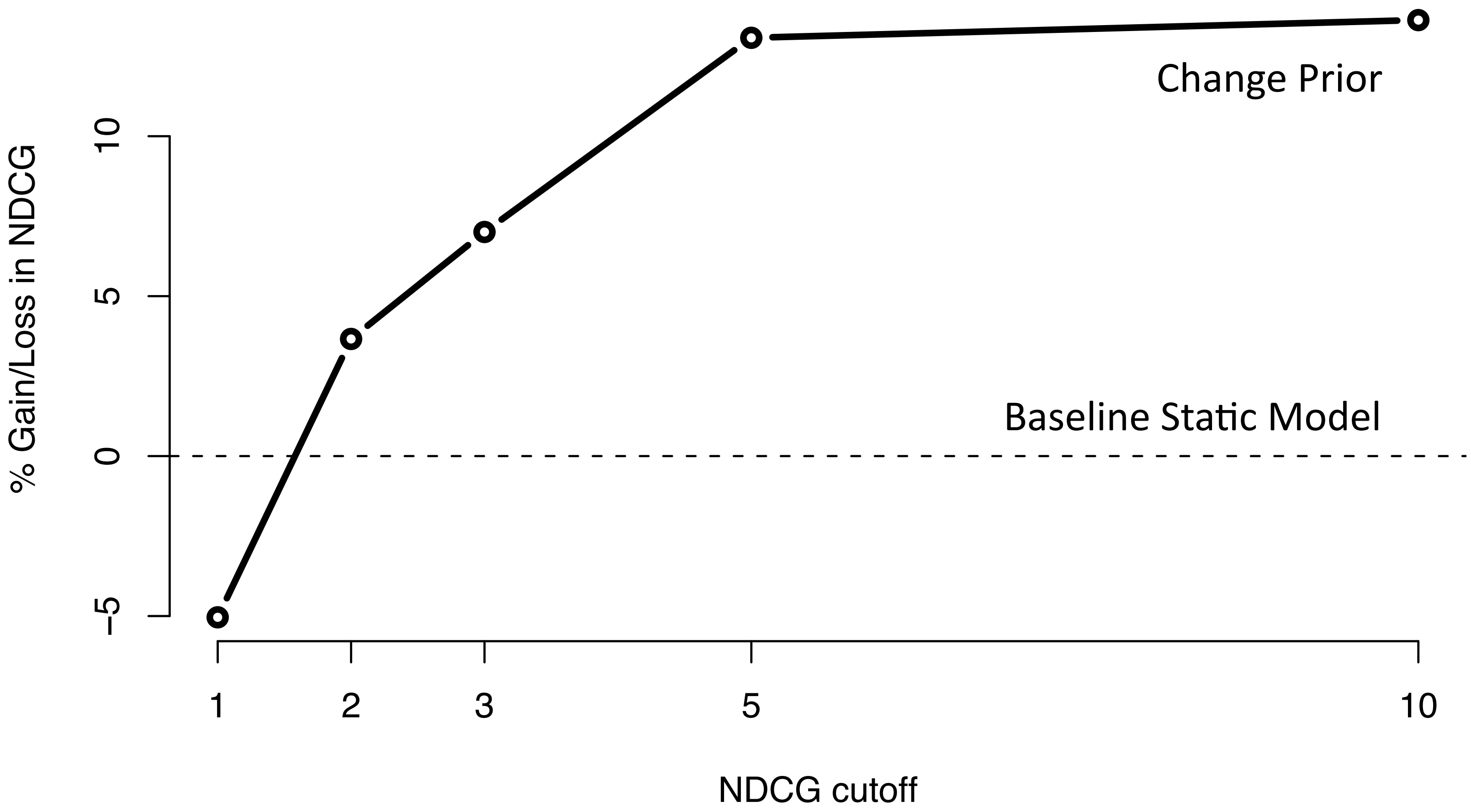
Uniform Prior:

$$P(D) \propto 1.0$$

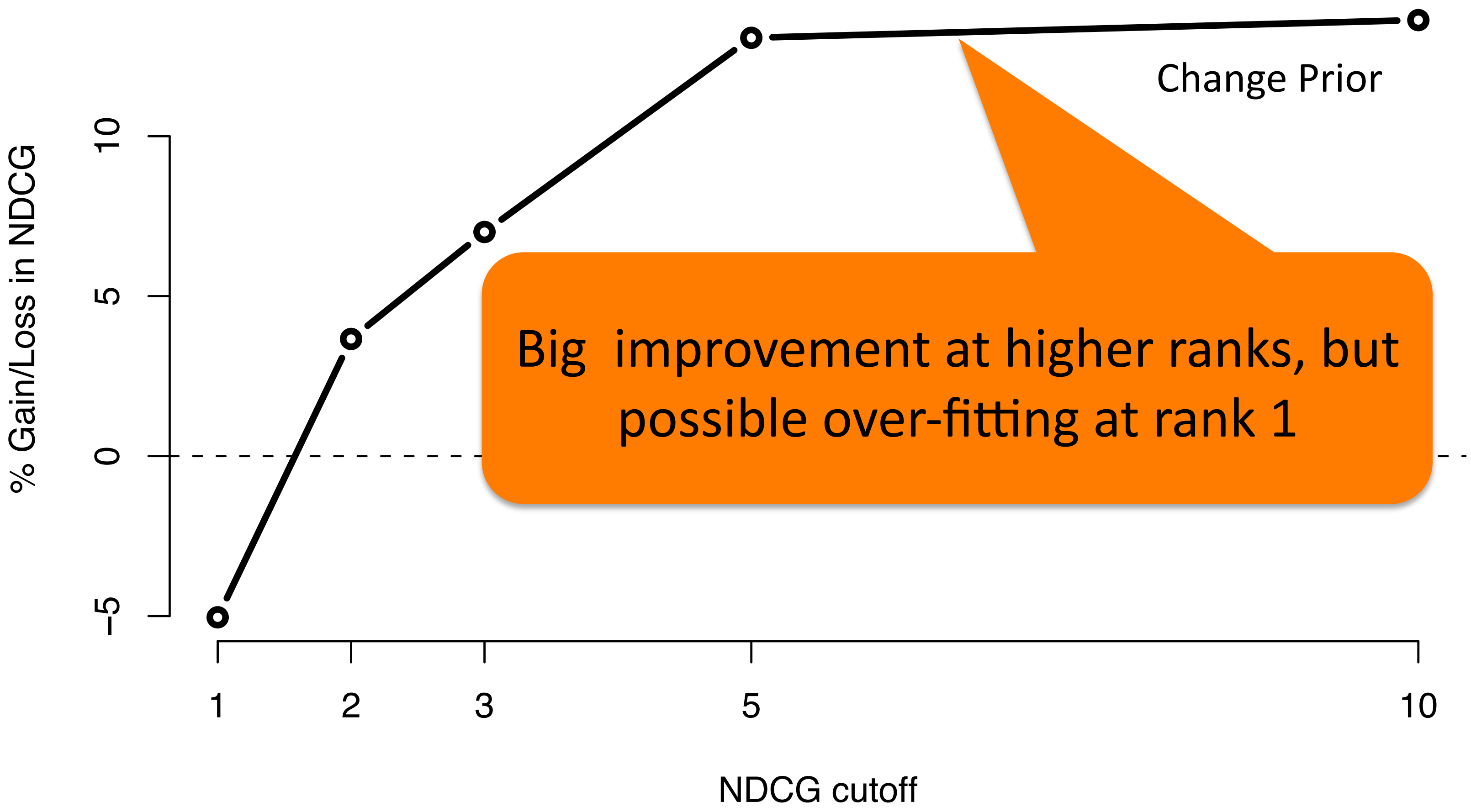
“Change” Prior:

$$P_{ch}(D) \propto (ShDiff(D) + 1)^{\gamma}$$

# Favoring Dynamic Documents



# Favoring Dynamic Documents

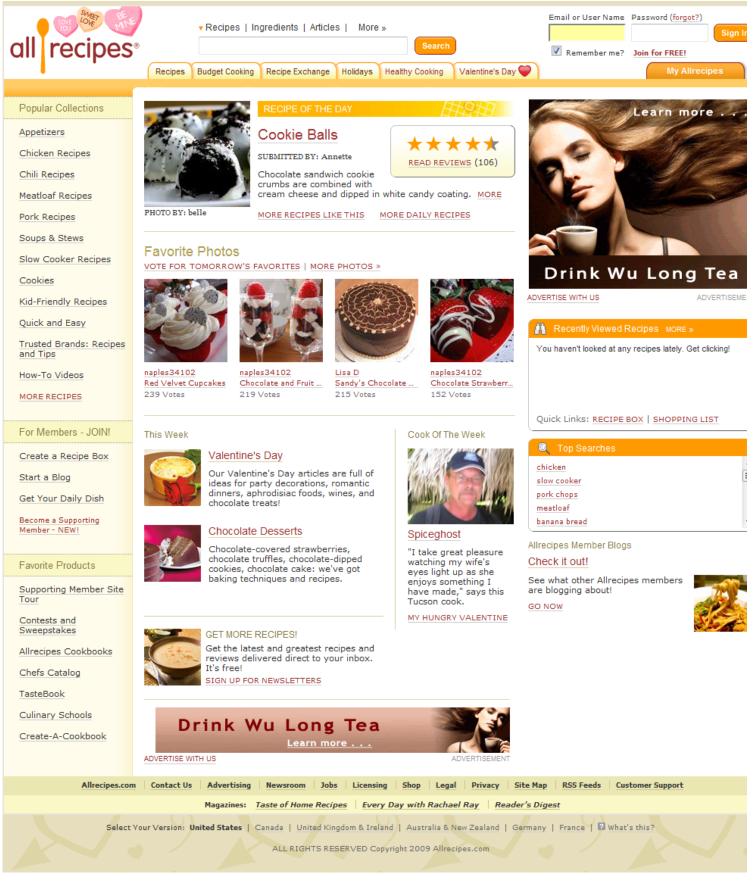


Big improvement at higher ranks, but possible over-fitting at rank 1

# Change Within the Document

Are there characteristics of a document’s dynamic content that indicate some content may be more important?

# Change Within the Document



Recipes | Ingredients | Articles | More »

Search

Email or User Name

Password (forgot?)

Sign In

☑ Remember me? Join for FREE!

Recipes | Budget Cooking | Recipe Exchange | Holidays | Healthy Cooking | Valentine's Day

My Allrecipes

Popular Collections

- Appetizers
- Chicken Recipes
- Chili Recipes
- Meatloaf Recipes
- Pork Recipes
- Soups & Stews
- Slow Cooker Recipes
- Cookies
- Kid-Friendly Recipes
- Quick and Easy
- Trusted Brands: Recipes and Tips
- How-To Videos
- MORE RECIPES

For Members - JOIN!

- Create a Recipe Box
- Start a Blog
- Get Your Daily Dish
- Become a Supporting Member - NEW!

Favorite Products

- Supporting Member Site Tour
- Contests and Sweepstakes
- Allrecipes Cookbooks
- Chefs Catalog
- TasteBook
- Culinary Schools
- Create-A-Cookbook

RECIPE OF THE DAY

Cookie Balls

SUBMITTED BY: Annette

★★★★★ READ REVIEWS (106)

Chocolate sandwich cookie crumbs are combined with cream cheese and dipped in white candy coating. MORE

PHOTO BY: belle

MORE RECIPES LIKE THIS MORE DAILY RECIPES

Favorite Photos

VOTE FOR TOMORROW'S FAVORITES | MORE PHOTOS »

| naples34102<br>Red Velvet Cupcakes<br>239 Votes | naples34102<br>Chocolate and Fruit ...<br>219 Votes | Lisa D<br>Sandy's Chocolate ...<br>215 Votes | naples34102<br>Chocolate Strawberr...<br>152 Votes |
|---|---|---|---|

This Week

Valentine's Day

Our Valentine's Day articles are full of ideas for party decorations, romantic dinners, aphrodisiac foods, wines, and chocolate treats!

Chocolate Desserts

Chocolate-covered strawberries, chocolate truffles, chocolate-dipped cookies, chocolate cake: we've got baking techniques and recipes.

GET MORE RECIPES!

Get the latest and greatest recipes and reviews delivered direct to your inbox. It's free!

SIGN UP FOR NEWSLETTERS

Cook Of The Week

Spiceghost

"I take great pleasure watching my wife's eyes light up as she enjoys something I have made," says this Tucson cook.

MY HUNGRY VALENTINE

Recently Viewed Recipes MORE »

You haven't looked at any recipes lately. Get clicking!

Quick Links: RECIPE BOX | SHOPPING LIST

Top Searches

- chicken
- slow cooker
- pork chops
- meatloaf
- banana bread

Allrecipes Member Blogs

Check it out!

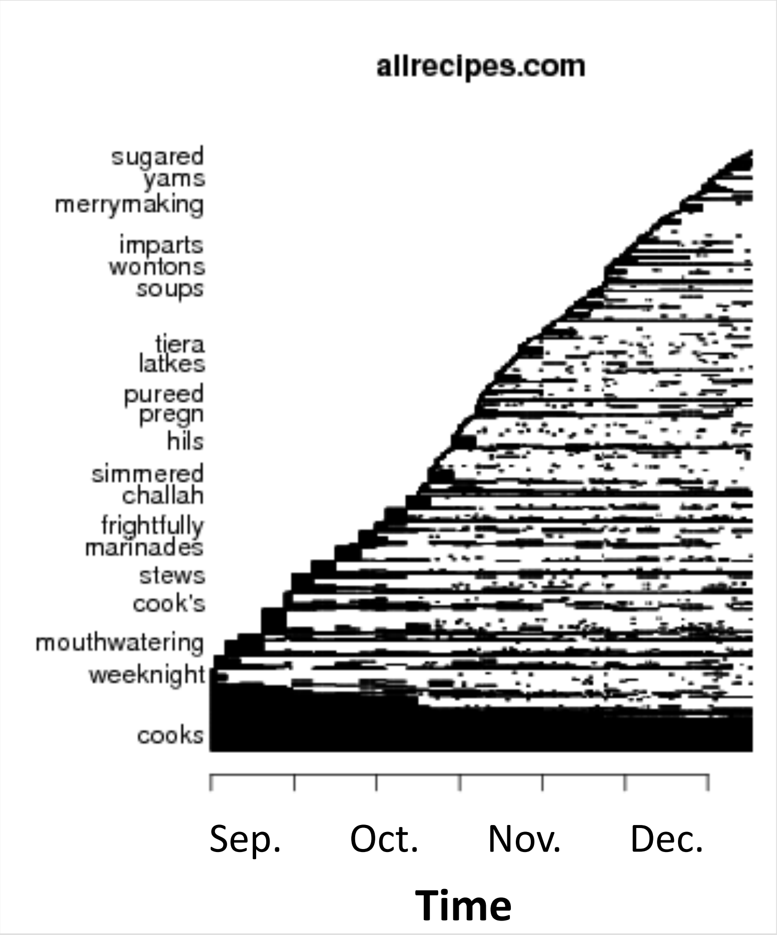
See what other Allrecipes members are blogging about!

GO NOW

Allrecipes.com | Contact Us | Advertising | Newsroom | Jobs | Licensing | Shop | Legal | Privacy | Site Map | RSS Feeds | Customer Support

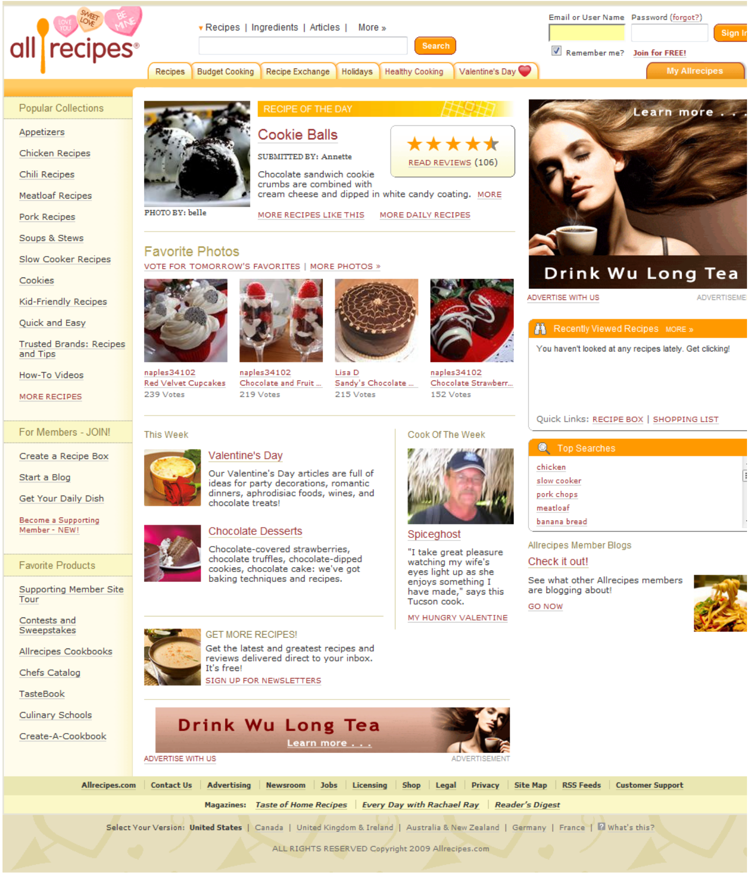
Magazines: *Taste of Home Recipes* | *Every Day with Rachael Ray* | *Reader's Digest*

Select Your Version: **United States** | Canada | United Kingdom & Ireland | Australia & New Zealand | Germany | France | What's this?



Adar, Teevan, Dumais & Elsas, "The Web Changes Everything: Understanding the Dynamics of Web Content" WSDM 2009

# Change Within the Document



allrecipes

▾ Recipes | Ingredients | Articles | More »

Search

Email or User Name Password (forgot?) Sign In

☑ Remember me? Join for FREE!

Recipes | Budget Cooking | Recipe Exchange | Holidays | Healthy Cooking | Valentine's Day | My Allrecipes

Popular Collections

- Appetizers
- Chicken Recipes
- Chili Recipes
- Meatloaf Recipes
- Pork Recipes
- Soups & Stews
- Slow Cooker Recipes
- Cookies
- Kid-Friendly Recipes
- Quick and Easy
- Trusted Brands: Recipes and Tips
- How-To Videos
- MORE RECIPES

For Members - JOIN!

- Create a Recipe Box
- Start a Blog
- Get Your Daily Dish
- Become a Supporting Member - NEW!

Favorite Products

- Supporting Member Site Tour
- Contests and Sweepstakes
- Allrecipes Cookbooks
- Chefs Catalog
- TasteBook
- Culinary Schools
- Create-A-Cookbook

PHOTO BY: belle

RECIPE OF THE DAY

Cookie Balls

SUBMITTED BY: Annette

READ REVIEWS (106)

Chocolate sandwich cookie crumbs are combined with cream cheese and dipped in white candy coating. MORE

MORE RECIPES LIKE THIS MORE DAILY RECIPES

Favorite Photos

VOTE FOR TOMORROW'S FAVORITES | MORE PHOTOS »

naples34102
Red Velvet Cupcakes
239 Votes

naples34102
Chocolate and Fruit ...
219 Votes

Lisa D
Sandy's Chocolate ...
215 Votes

naples34102
Chocolate Strawberr...
152 Votes

This Week

Valentine's Day

Our Valentine's Day articles are full of ideas for party decorations, romantic dinners, aphrodisiac foods, wines, and chocolate treats!

Chocolate Desserts

Chocolate-covered strawberries, chocolate truffles, chocolate-dipped cookies, chocolate cake: we've got baking techniques and recipes.

Cook Of The Week

Spiceghost

"I take great pleasure watching my wife's eyes light up as she enjoys something I have made," says this Tucson cook.

MY HUNGRY VALENTINE

GET MORE RECIPES!

Get the latest and greatest recipes and reviews delivered direct to your inbox. It's free!

SIGN UP FOR NEWSLETTERS

Recently Viewed Recipes MORE »

You haven't looked at any recipes lately. Get clicking!

Quick Links: RECIPE BOX | SHOPPING LIST

Top Searches

- chicken
- slow cooker
- pork chops
- meatloaf
- banana bread

Allrecipes Member Blogs

Check it out!

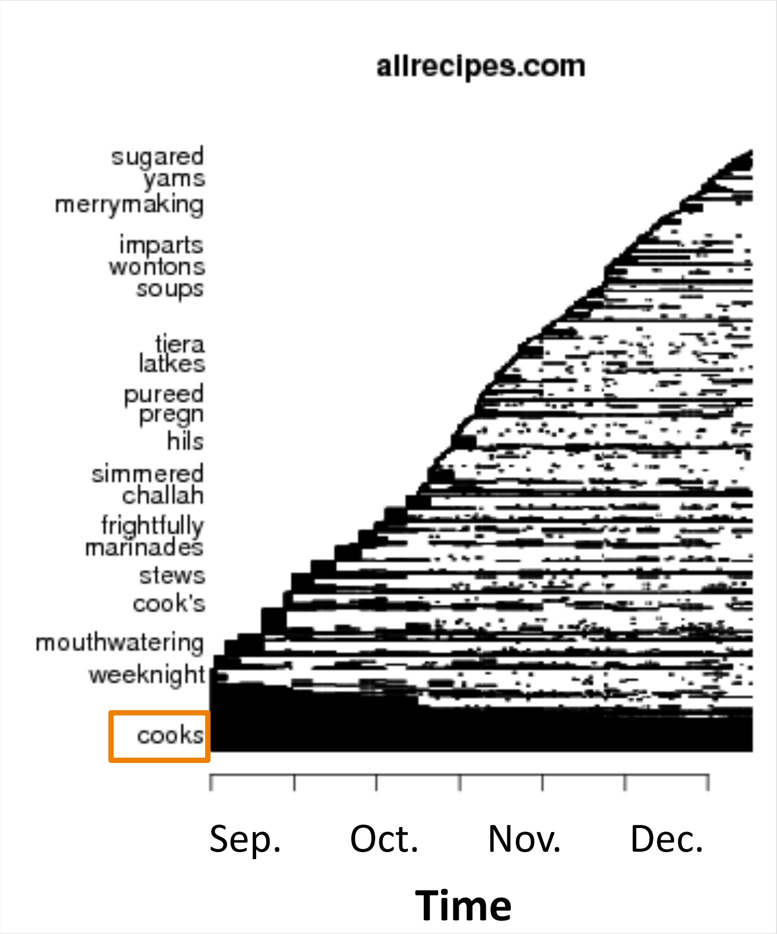
See what other Allrecipes members are blogging about!

GO NOW

Allrecipes.com | Contact Us | Advertising | Newsroom | Jobs | Licensing | Shop | Legal | Privacy | Site Map | RSS Feeds | Customer Support

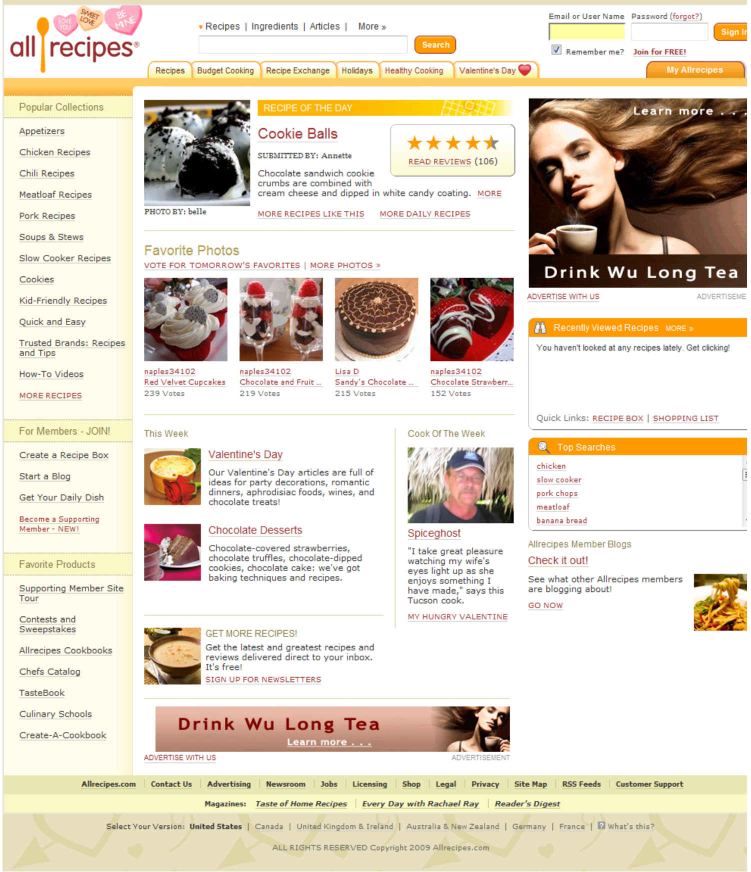
Magazines: *Taste of Home Recipes* | *Every Day with Rachael Ray* | *Reader's Digest*

Select Your Version: **United States** | Canada | United Kingdom & Ireland | Australia & New Zealand | Germany | France | What's this?



Adar, Teevan, Dumais & Elsas, "The Web Changes Everything: Understanding the Dynamics of Web Content" WSDM 2009

# Change Within the Document



Cookie Balls

SUBMITTED BY: Annette

Chocolate sandwich cookie crumbs are combined with cream cheese and dipped in white candy coating.

PHOTO BY: belle

Favorite Photos

| | | | |
|---|---|---|---|
| naples34102 | naples34102 | Lisa D | naples34102 |
| Red Velvet Cupcakes | Chocolate and Fruit ... | Sandy's Chocolate ... | Chocolate Strawberr... |
| 239 Votes | 219 Votes | 215 Votes | 152 Votes |

You haven't looked at any recipes lately. Get clicking!

This Week

Valentine's Day

Our Valentine's Day articles are full of ideas for party decorations, romantic dinners, aphrodisiac foods, wines, and chocolate treats!

Chocolate Desserts

Chocolate-covered strawberries, chocolate truffles, chocolate-dipped cookies, chocolate cake: we've got baking techniques and recipes.

Cook Of The Week

Spiceghost

"I take great pleasure watching my wife's eyes light up as she enjoys something I have made," says this Tucson cook.

Allrecipes Member Blogs

Check it out!

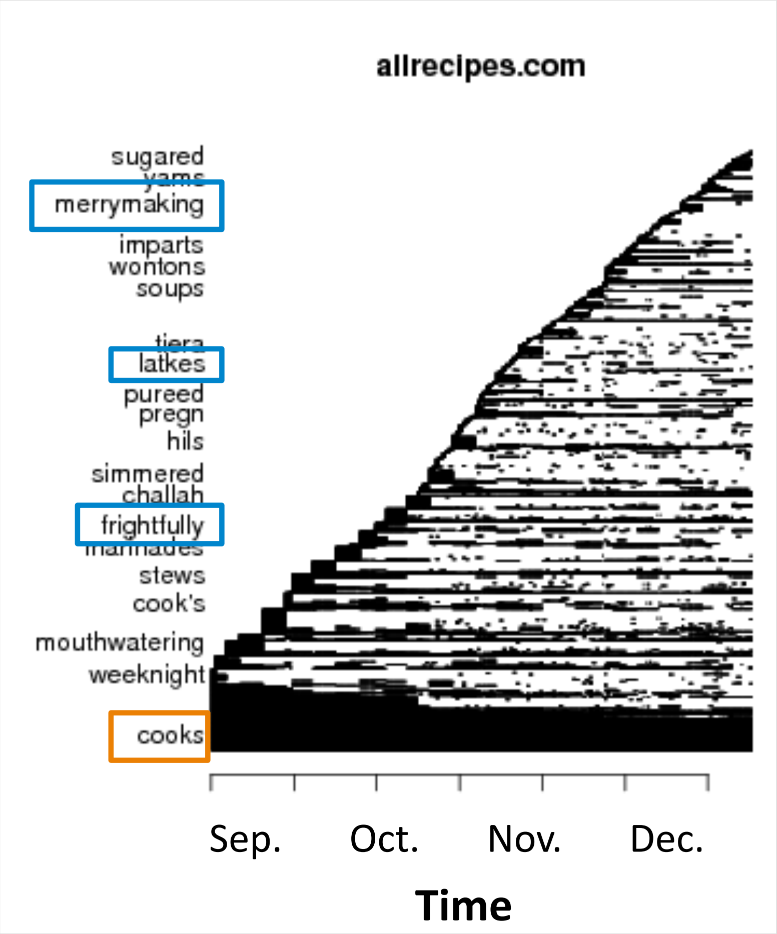
See what other Allrecipes members are blogging about!

GET MORE RECIPES!

Get the latest and greatest recipes and reviews delivered direct to your inbox. It's free!

Magazines: *Taste of Home Recipes* | *Every Day with Rachael Ray* | *Reader's Digest*

ALL RIGHTS RESERVED Copyright 2009 Allrecipes.com

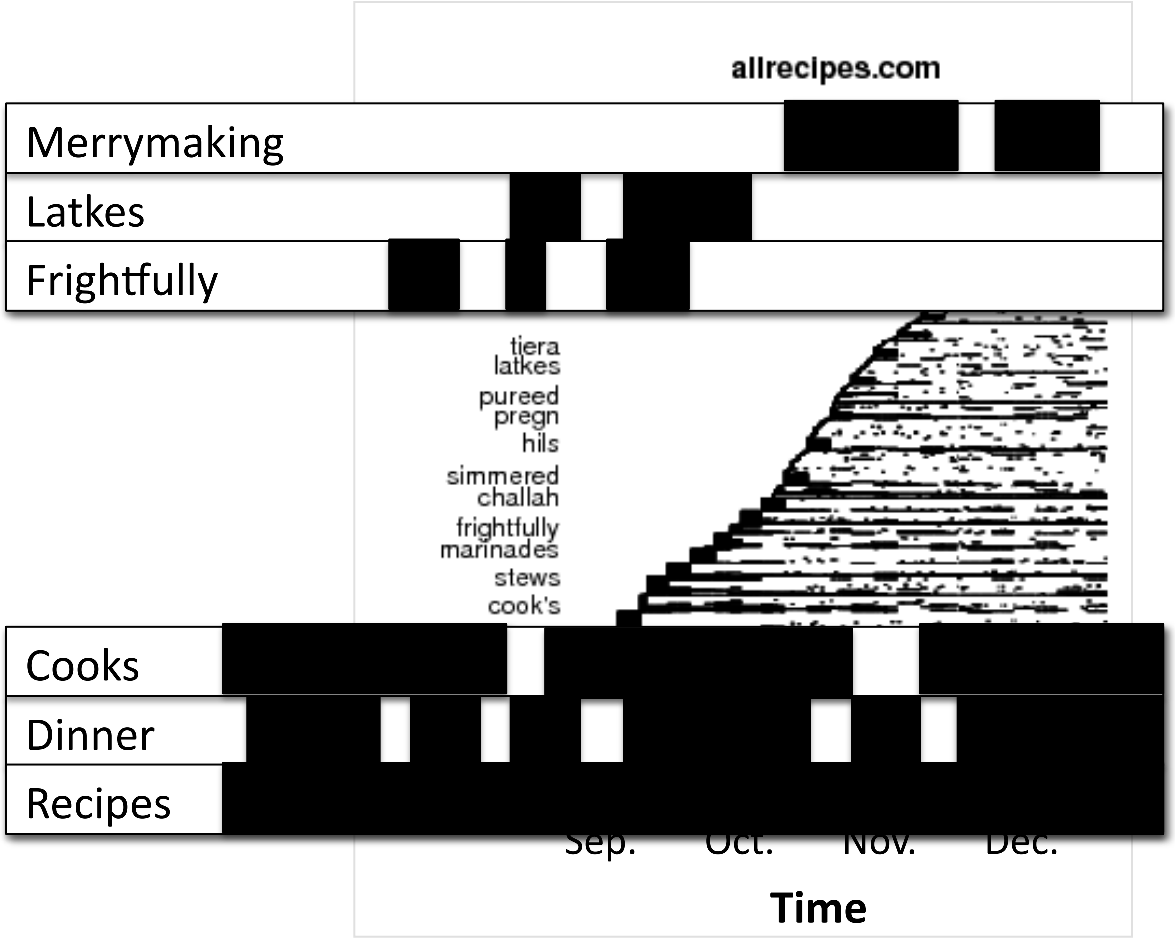


Adar, Teevan, Dumais & Elsas, “The Web Changes Everything: Understanding the Dynamics of Web Content” WSDM 2009

# Change Within the Document

Short-Lived Vocabulary

Long-Lived Vocabulary



# Leveraging Change Within the Document

Identifying transient & permanent vocabulary:

- **Short-lived**: come & go quickly

  in fewer than 50% of the document's slices
- **Medium-lived**:

  in 50-90% of the document's slices
- **Long-lived**: tend to stick for a long time

  in > 90% of the document's slices

# Leveraging Change Within the Document

Model relevance as a *mixture* of LONG- MEDIUM- and SHORT-lived vocabulary:

$$
\begin{aligned}
P(D|Q) \propto P(D)\Big(&\lambda_L P(Q|D_L) \\
&+\lambda_M P(Q|D_M) \\
&+\lambda_S P(Q|D_S)\Big)
\end{aligned}
$$

“Dynamic Model”

# Leveraging Change Within the Document

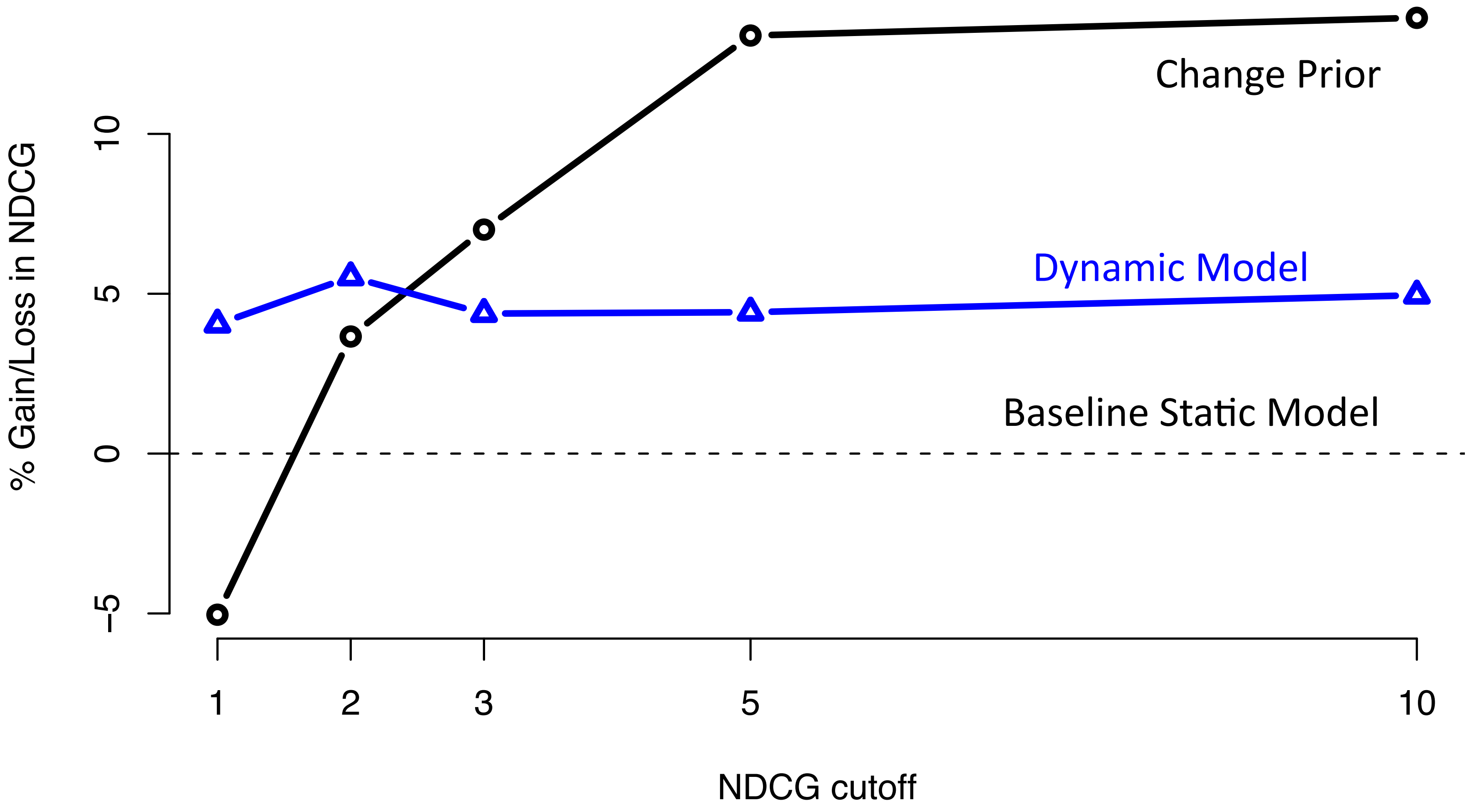
Model relevance as a *mixture* of LONG-MEDIUM- and SHORT-lived vocabulary:

$$
\begin{aligned}
P(D|Q) \propto P(D)\Big(& \boxed{\lambda_L P(Q|D_L)} \\
&+\lambda_M P(Q|D_M) \\
&+\boxed{\lambda_S P(Q|D_S)}\Big)
\end{aligned}
$$

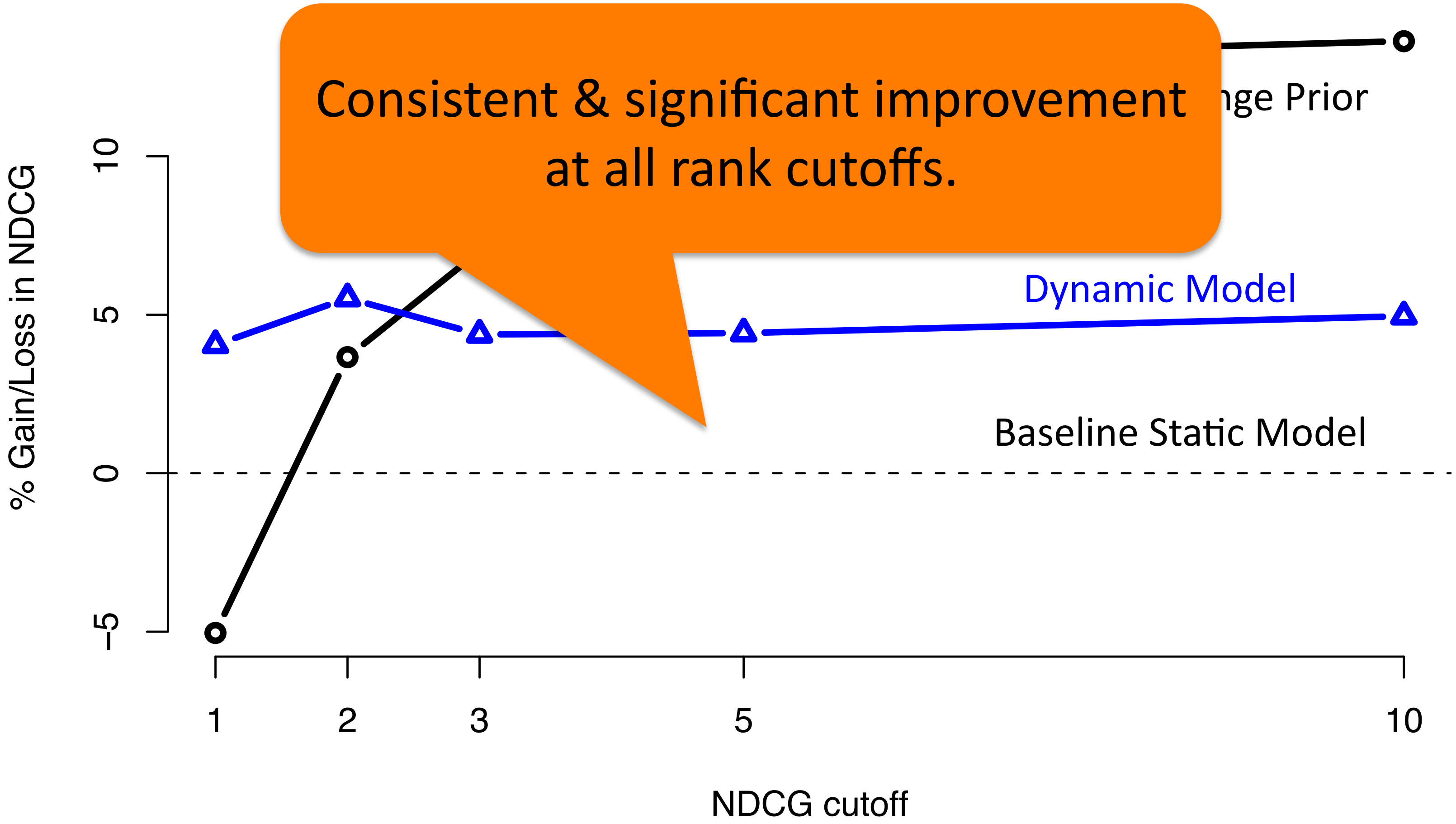
Differentially weight long-lived and short-lived vocabulary.

“Dynamic Model”

# Leveraging Change Within the Document



# Leveraging Change Within the Document

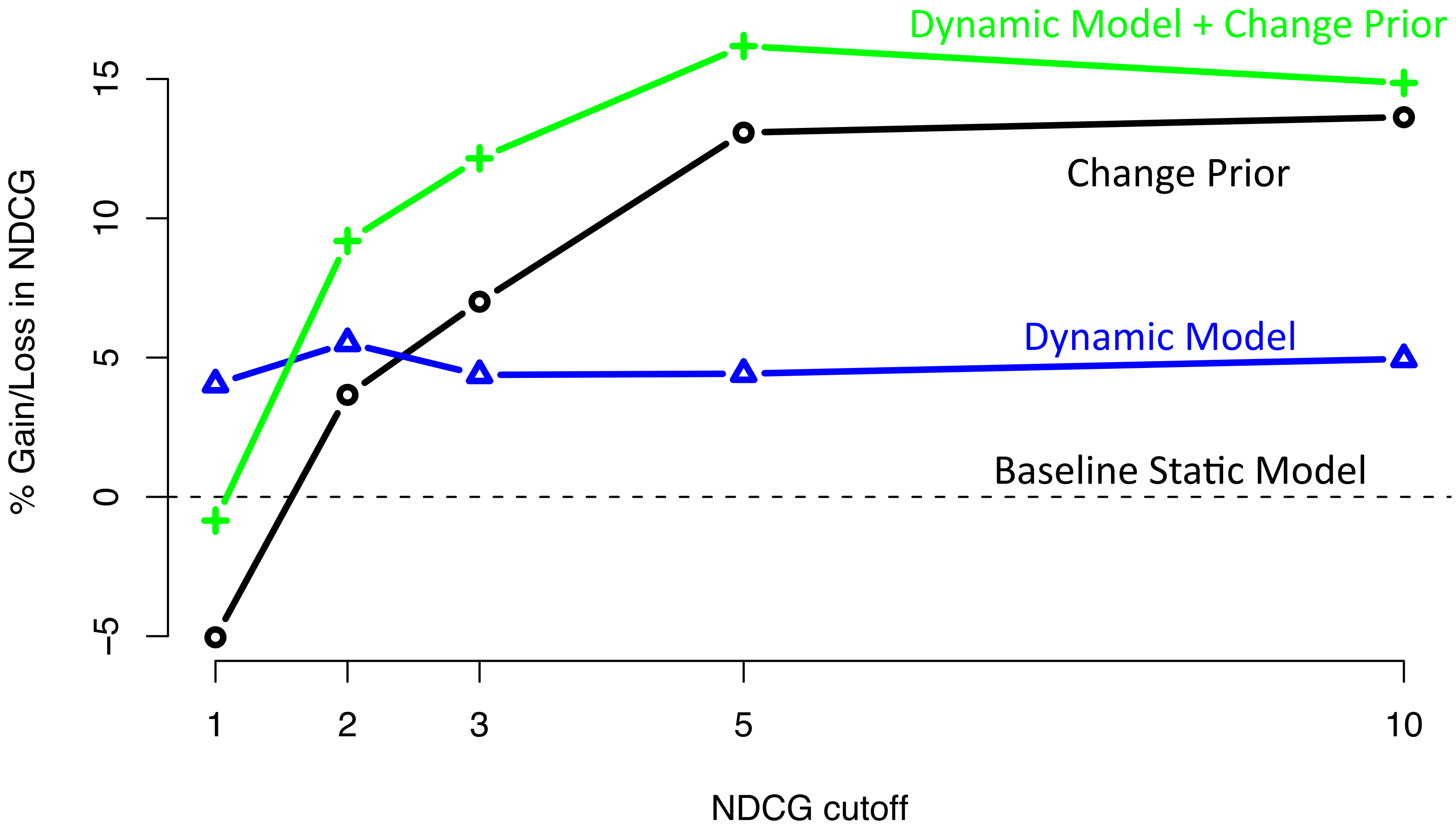


Consistent & significant improvement at all rank cutoffs.

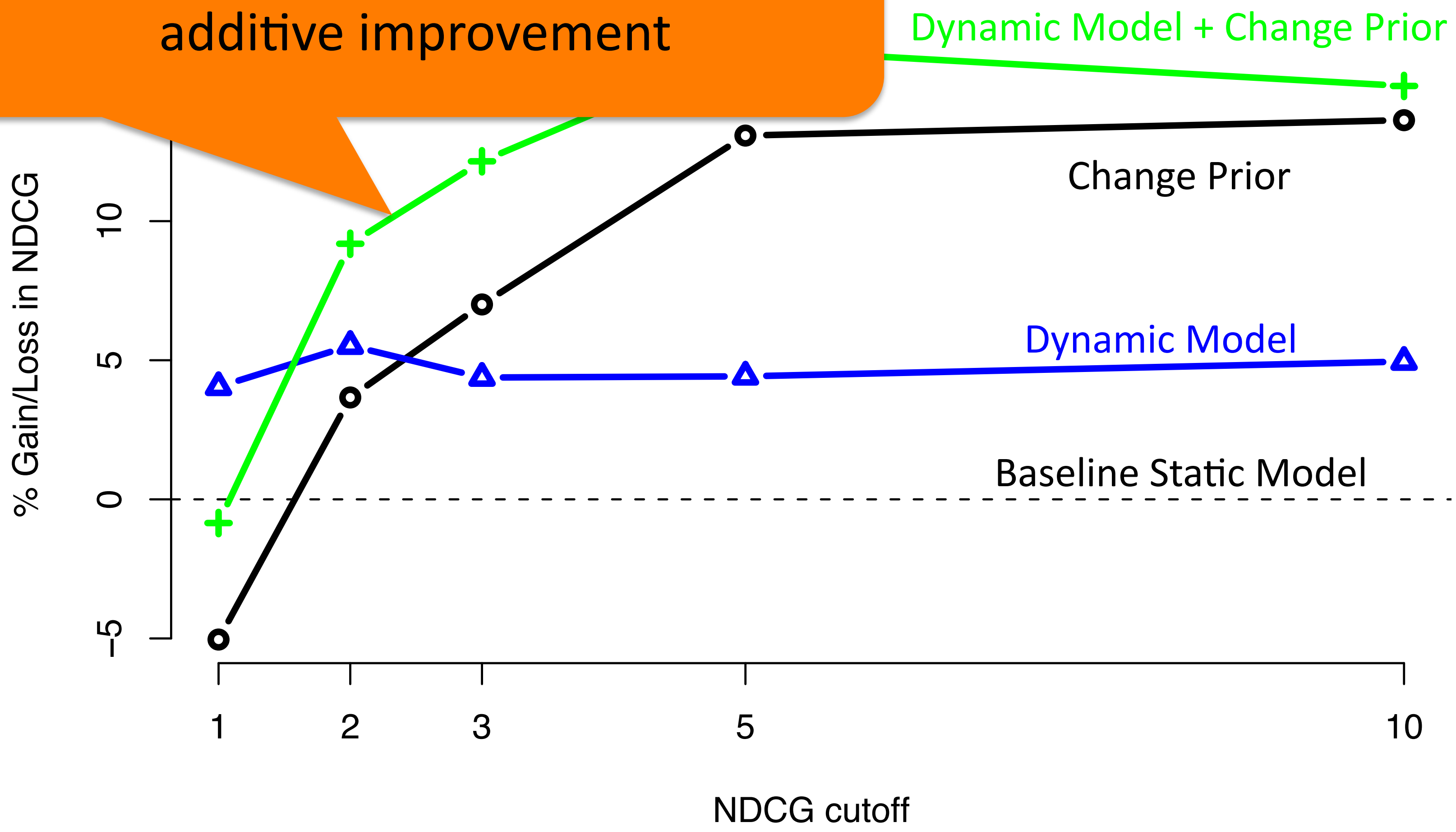
# Change & Relevance Ranking

- Presented two methods for leveraging changing content in relevance ranking:
  - **Query-Independent Change Prior,** favoring dynamic documents irrespective of query
  - **Dynamic Document Representation**, differentially weighting long-term and short-term vocabulary
- Combined Model: Best of both worlds?

# Combined Model



# Combined Model



Combined model shows (roughly) additive improvement

# Conclusion & Next Steps

- Documents change, and we can use characteristics of those dynamics to improve retrieval performance.
- Presented two complementary methods of leveraging change in ranking.
- Focus here on navigational queries; current work is looking at *dynamic* information needs.

Relevance may change over time.

Thank You

Questions?