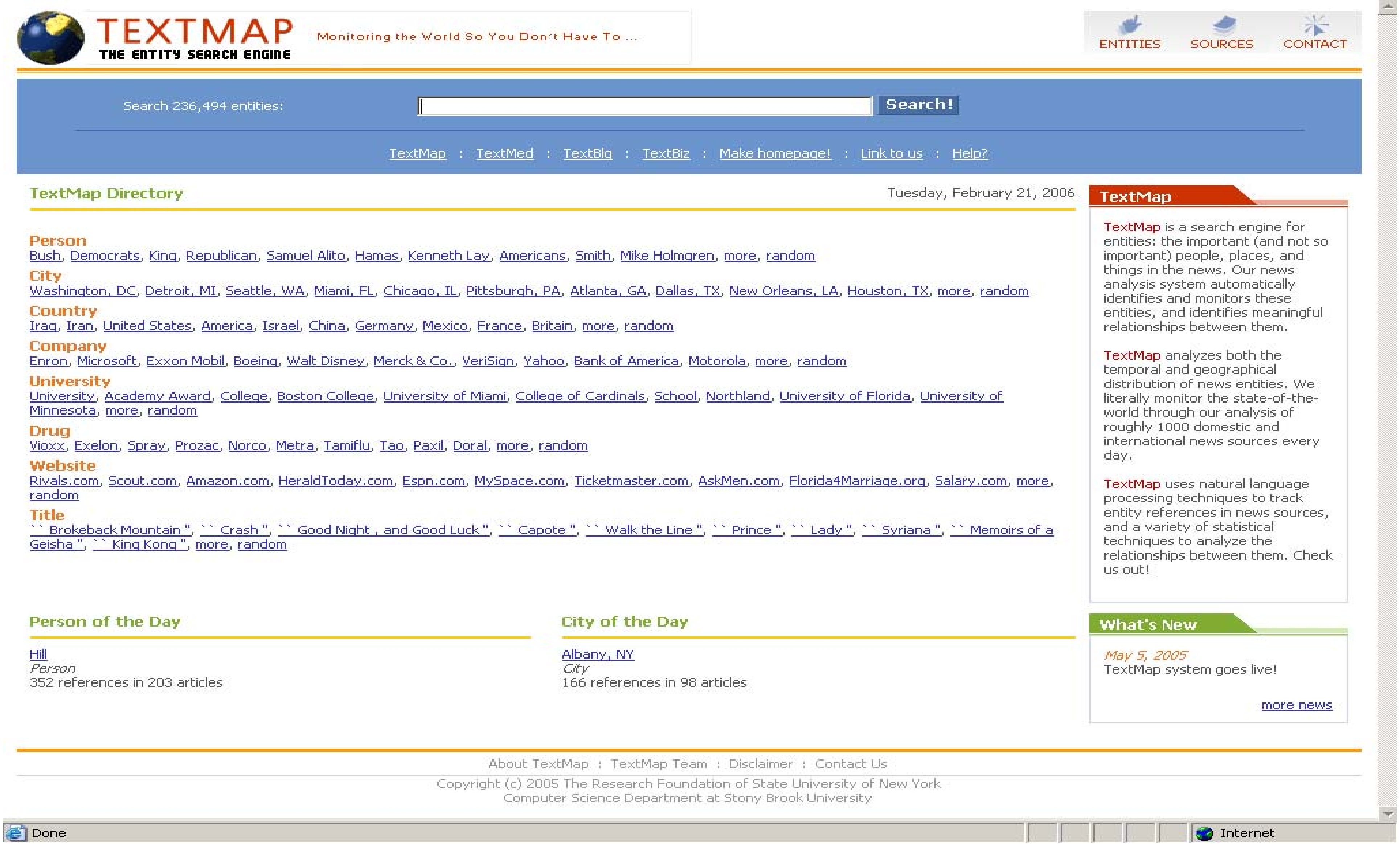
News and Blog Analysis with Lydia

Steven Skiena
Dept. of Computer Science
SUNY Stony Brook
http://www.cs.sunysb.edu/~skiena

Large-Scale News Analysis

- Our Lydia news analysis system does a daily analysis of over 1000+ online English and foreign-language newspapers, plus blogs, RSS feeds, and other news sources.
- We currently track over 1,000,000 news entities, providing spatial, temporal, relational and sentiment analysis
- We believe our data and analysis should be of great interest in political science and related fields.

www.textmap.com


TEXTMAP
THE ENTITY SEARCH ENGINE
Monitoring the World So You Don't Have To ...
ENTITIES
SOURCES
CONTACT
Search!
TextMap : TextMed : TextBlg : TextBiz : Make homepage! : Link to us : Help?
Done
Internet

Search 236,494 entities:

TextMap Directory Tuesday, February 21, 2006

Person
Bush, Democrats, King, Republican, Samuel Alito, Hamas, Kenneth Lay, Americans, Smith, Mike Holmgren, more, random

City
Washington, DC, Detroit, MI, Seattle, WA, Miami, FL, Chicago, IL, Pittsburgh, PA, Atlanta, GA, Dallas, TX, New Orleans, LA, Houston, TX, more, random

Country
Iraq, Iran, United States, America, Israel, China, Germany, Mexico, France, Britain, more, random

Company
Enron, Microsoft, Exxon Mobil, Boeing, Walt Disney, Merck & Co., VeriSign, Yahoo, Bank of America, Motorola, more, random

University
University, Academy Award, College, Boston College, University of Miami, College of Cardinals, School, Northland, University of Florida, University of Minnesota, more, random

Drug
Vioxx, Exelon, Spray, Prozac, Norco, Metra, Tamiflu, Tao, Paxil, Doral, more, random

Website
Rivals.com, Scout.com, Amazon.com, HeraldToday.com, Espn.com, MySpace.com, Ticketmaster.com, AskMen.com, Florida4Marriage.org, Salary.com, more, random

Title
`` Brokeback Mountain '', `` Crash '', `` Good Night , and Good Luck '', `` Capote '', `` Walk the Line '', `` Prince '', `` Lady '', `` Syriana '', `` Memoirs of a Geisha '', `` King Kong '', more, random

Person of the Day

Hill
Person
352 references in 203 articles

City of the Day

Albany, NY
City
166 references in 98 articles

TextMap

TextMap is a search engine for entities: the important (and not so important) people, places, and things in the news. Our news analysis system automatically identifies and monitors these entities, and identifies meaningful relationships between them.

TextMap analyzes both the temporal and geographical distribution of news entities. We literally monitor the state-of-the-world through our analysis of roughly 1000 domestic and international news sources every day.

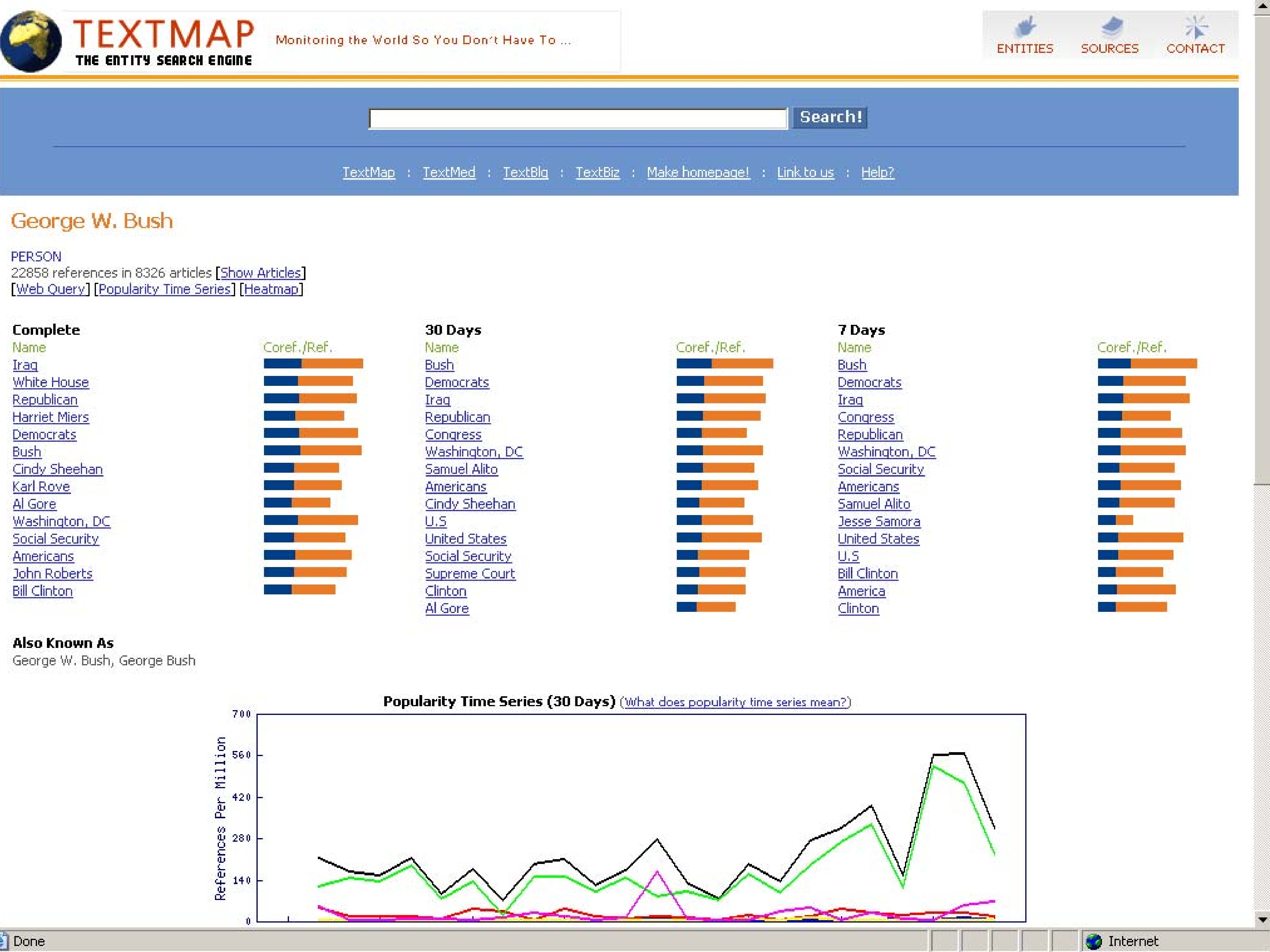
TextMap uses natural language processing techniques to track entity references in news sources, and a variety of statistical techniques to analyze the relationships between them. Check us out!

What's New

May 5, 2005
TextMap system goes live!

more news

About TextMap : TextMap Team : Disclaimer : Contact Us



TEXTMAP
THE ENTITY SEARCH ENGINE

Monitoring the World So You Don't Have To ...

TextMap : TextMed : TextBlg : TextBiz : Make homepage! : Link to us : Help?

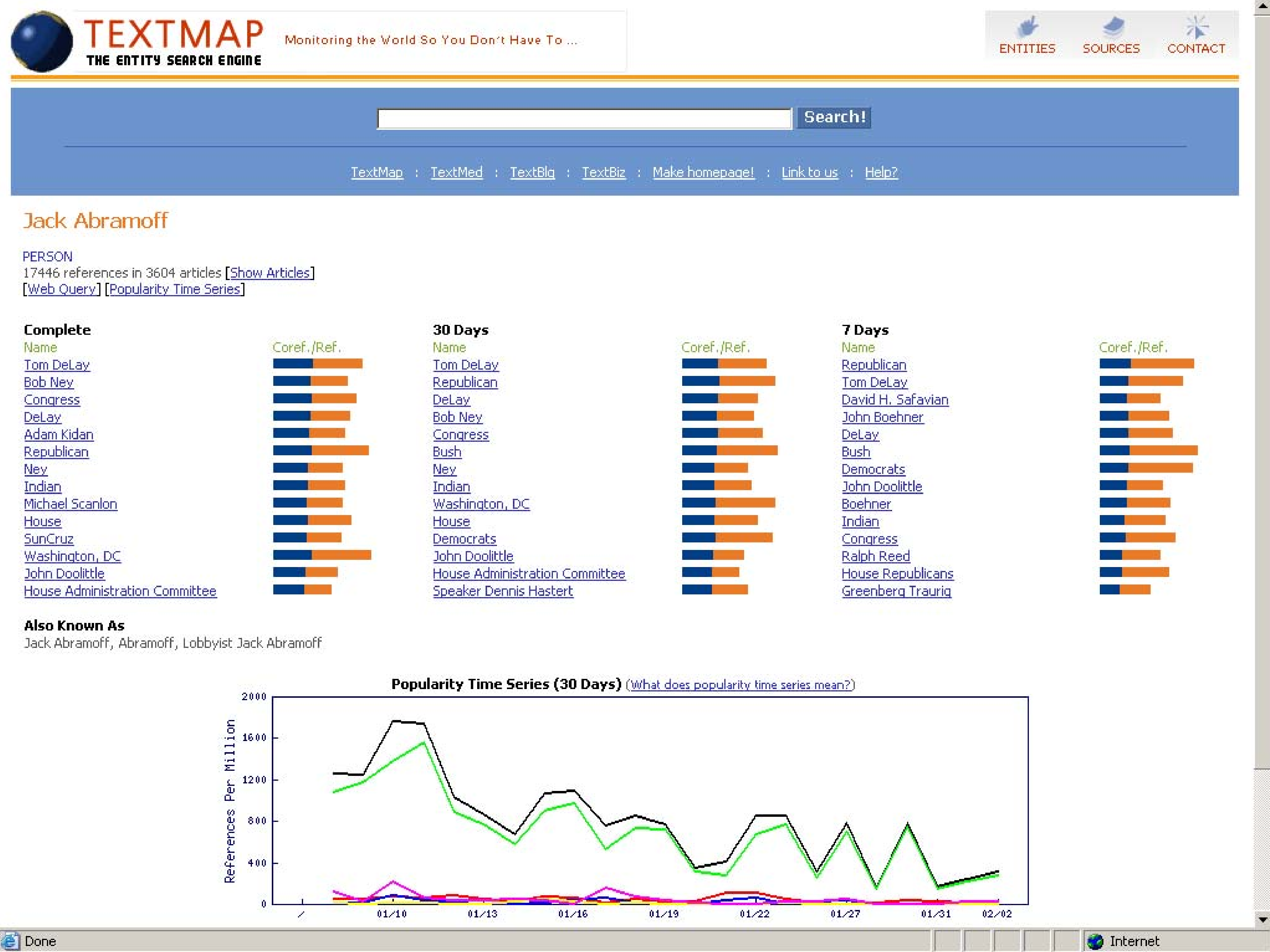
George W. Bush

PERSON
22858 references in 8326 articles [Show Articles]
[Web Query] [Popularity Time Series] [Heatmap]

Complete		30 Days		7 Days	
Name	Coref./Ref.	Name	Coref./Ref.	Name	Coref./Ref.
Iraq		Bush		Bush	
White House		Democrats		Democrats	
Republican		Iraq		Iraq	
Harriet Miers		Republican		Congress	
Democrats		Congress		Republican	
Bush		Washington, DC		Washington, DC	
Cindy Sheehan		Samuel Alito		Social Security	
Karl Rove		Americans		Americans	
Al Gore		Cindy Sheehan		Samuel Alito	
Washington, DC		U.S		Jesse Samora	
Social Security		United States		United States	
Americans		Social Security		U.S	
John Roberts		Supreme Court		Bill Clinton	
Bill Clinton		Clinton		America	
		Al Gore		Clinton	

Also Known As
George W. Bush, George Bush

Popularity Time Series (30 Days) (What does popularity time series mean?)



TEXTMAP
THE ENTITY SEARCH ENGINE
Monitoring the World So You Don't Have To ...
ENTITIES
SOURCES
CONTACT
Search!
TextMap : TextMed : TextBlg : TextBiz : Make homepage! : Link to us : Help?

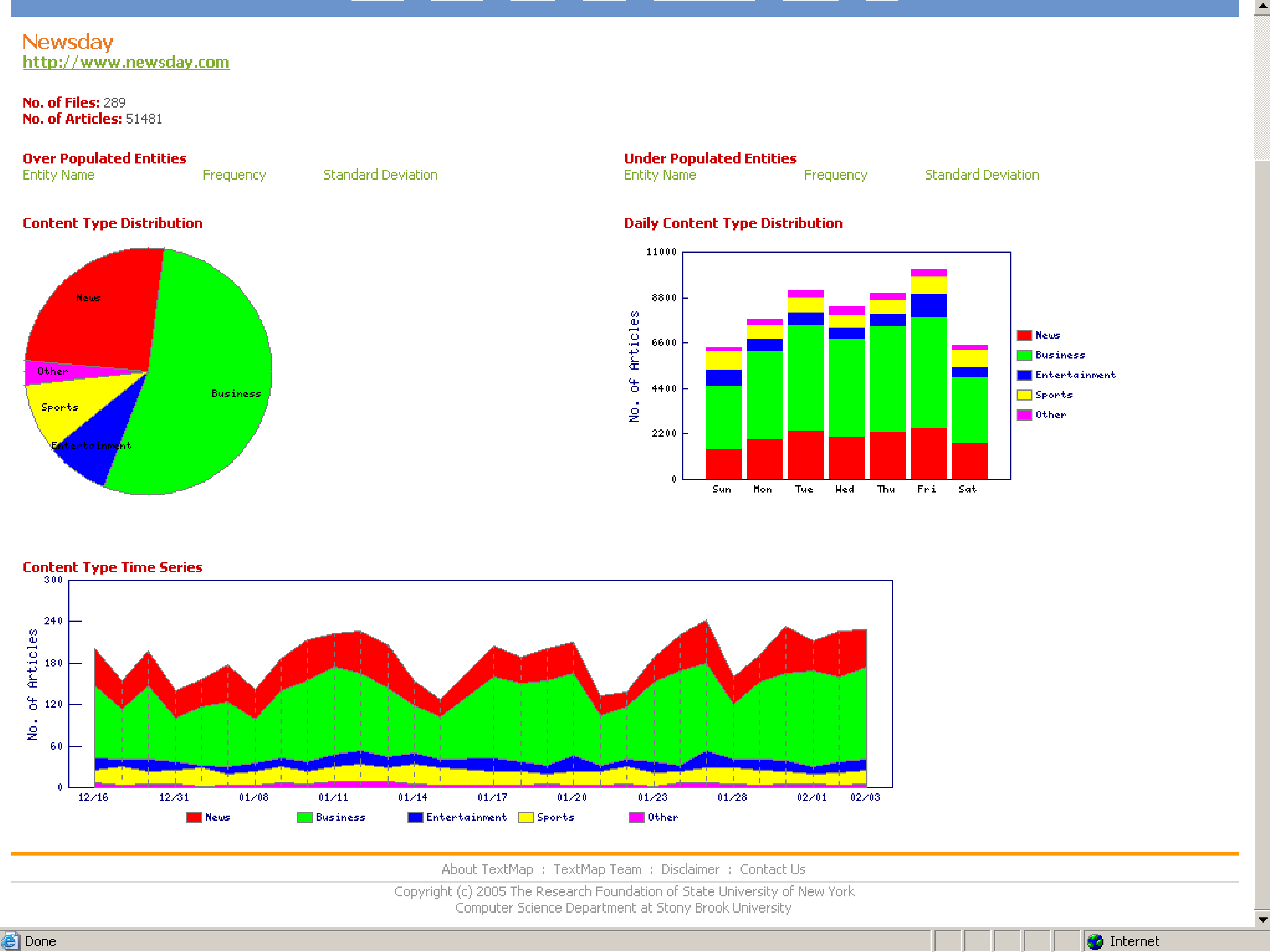
Jack Abramoff

PERSON
17446 references in 3604 articles [Show Articles]
[Web Query] [Popularity Time Series]

Complete		**30 Days**		**7 Days**	
Name	Coref./Ref.	Name	Coref./Ref.	Name	Coref./Ref.
Tom DeLay		Tom DeLay		Republican	
Bob Ney		Republican		Tom DeLay	
Congress		DeLay		David H. Safavian	
DeLay		Bob Ney		John Boehner	
Adam Kidan		Congress		DeLay	
Republican		Bush		Bush	
Ney		Ney		Democrats	
Indian		Indian		John Doolittle	
Michael Scanlon		Washington, DC		Boehner	
House		House		Indian	
SunCruz		Democrats		Congress	
Washington, DC		John Doolittle		Ralph Reed	
John Doolittle		House Administration Committee		House Republicans	
House Administration Committee		Speaker Dennis Hastert		Greenberg Traurig	

Also Known As
Jack Abramoff, Abramoff, Lobbyist Jack Abramoff

Popularity Time Series (30 Days) (What does popularity time series mean?)



Newsday

http://www.newsday.com

No. of Files: 289
No. of Articles: 51481

Over Populated Entities

Entity Name	Frequency	Standard Deviation

Under Populated Entities

Entity Name	Frequency	Standard Deviation

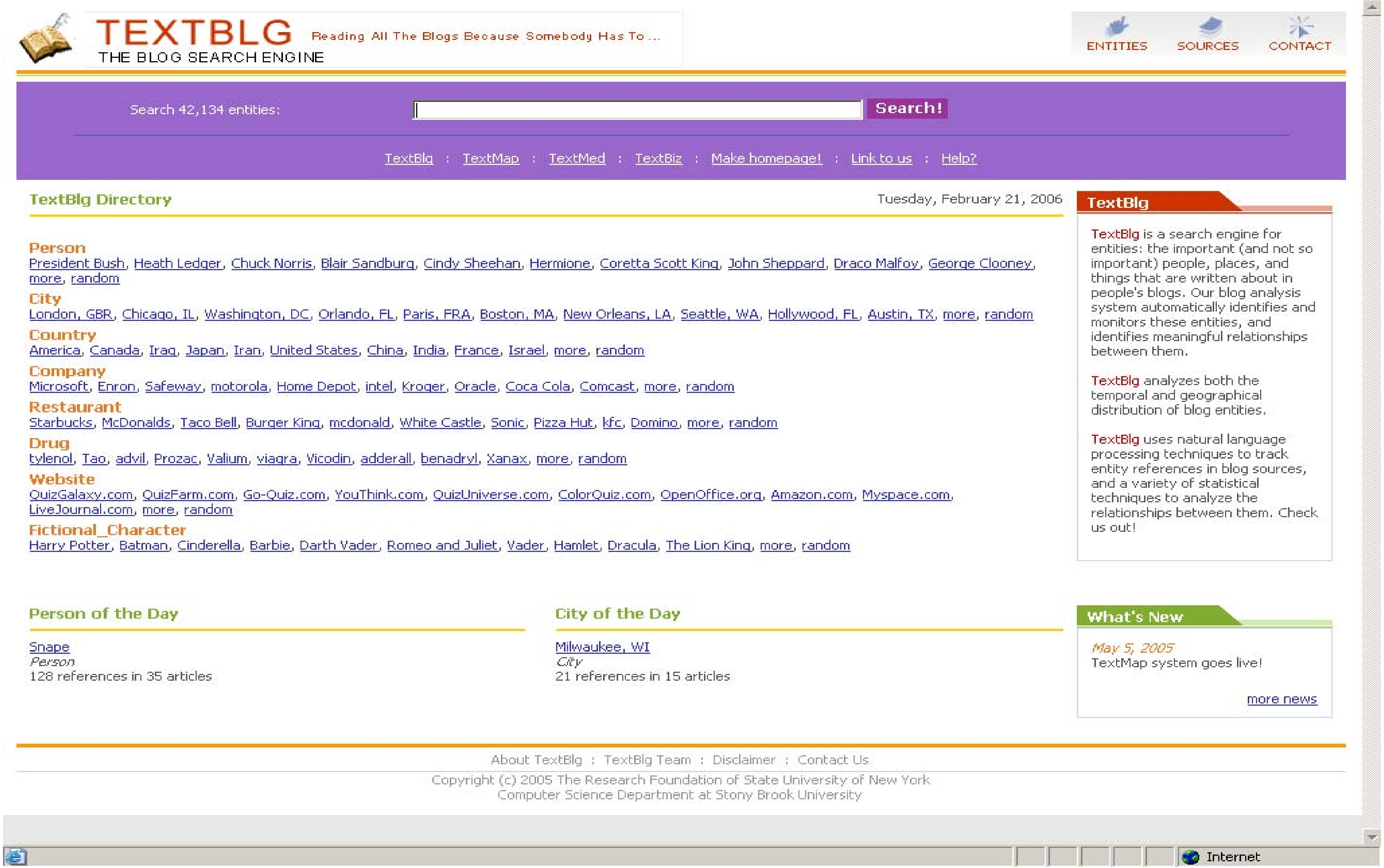
Content Type Distribution

Daily Content Type Distribution

Content Type Time Series

About TextMap : TextMap Team : Disclaimer : Contact Us

www.textblg.com



TEXTBLG
Reading All The Blogs Because Somebody Has To ...
THE BLOG SEARCH ENGINE
ENTITIES
SOURCES
CONTACT
Search!
TextBlg : TextMap : TextMed : TextBiz : Make homepage! : Link to us : Help?
Internet

Search 42,134 entities:

TextBlg Directory

Tuesday, February 21, 2006

Person
President Bush, Heath Ledger, Chuck Norris, Blair Sandburg, Cindy Sheehan, Hermione, Coretta Scott King, John Sheppard, Draco Malfoy, George Clooney, more, random

City
London, GBR, Chicago, IL, Washington, DC, Orlando, FL, Paris, FRA, Boston, MA, New Orleans, LA, Seattle, WA, Hollywood, FL, Austin, TX, more, random

Country
America, Canada, Iraq, Japan, Iran, United States, China, India, France, Israel, more, random

Company
Microsoft, Enron, Safeway, motorola, Home Depot, intel, Kroger, Oracle, Coca Cola, Comcast, more, random

Restaurant
Starbucks, McDonalds, Taco Bell, Burger King, mcdonald, White Castle, Sonic, Pizza Hut, kfc, Domino, more, random

Drug
tylenol, Tao, advil, Prozac, Valium, viagra, Vicodin, adderall, benadryl, Xanax, more, random

Website
QuizGalaxy.com, QuizFarm.com, Go-Quiz.com, YouThink.com, QuizUniverse.com, ColorQuiz.com, OpenOffice.org, Amazon.com, Myspace.com, LiveJournal.com, more, random

Fictional_Character
Harry Potter, Batman, Cinderella, Barbie, Darth Vader, Romeo and Juliet, Vader, Hamlet, Dracula, The Lion King, more, random

TextBlg

TextBlg is a search engine for entities: the important (and not so important) people, places, and things that are written about in people's blogs. Our blog analysis system automatically identifies and monitors these entities, and identifies meaningful relationships between them.

TextBlg analyzes both the temporal and geographical distribution of blog entities.

TextBlg uses natural language processing techniques to track entity references in blog sources, and a variety of statistical techniques to analyze the relationships between them. Check us out!

Person of the Day

Snape
Person
128 references in 35 articles

City of the Day

Milwaukee, WI
City
21 references in 15 articles

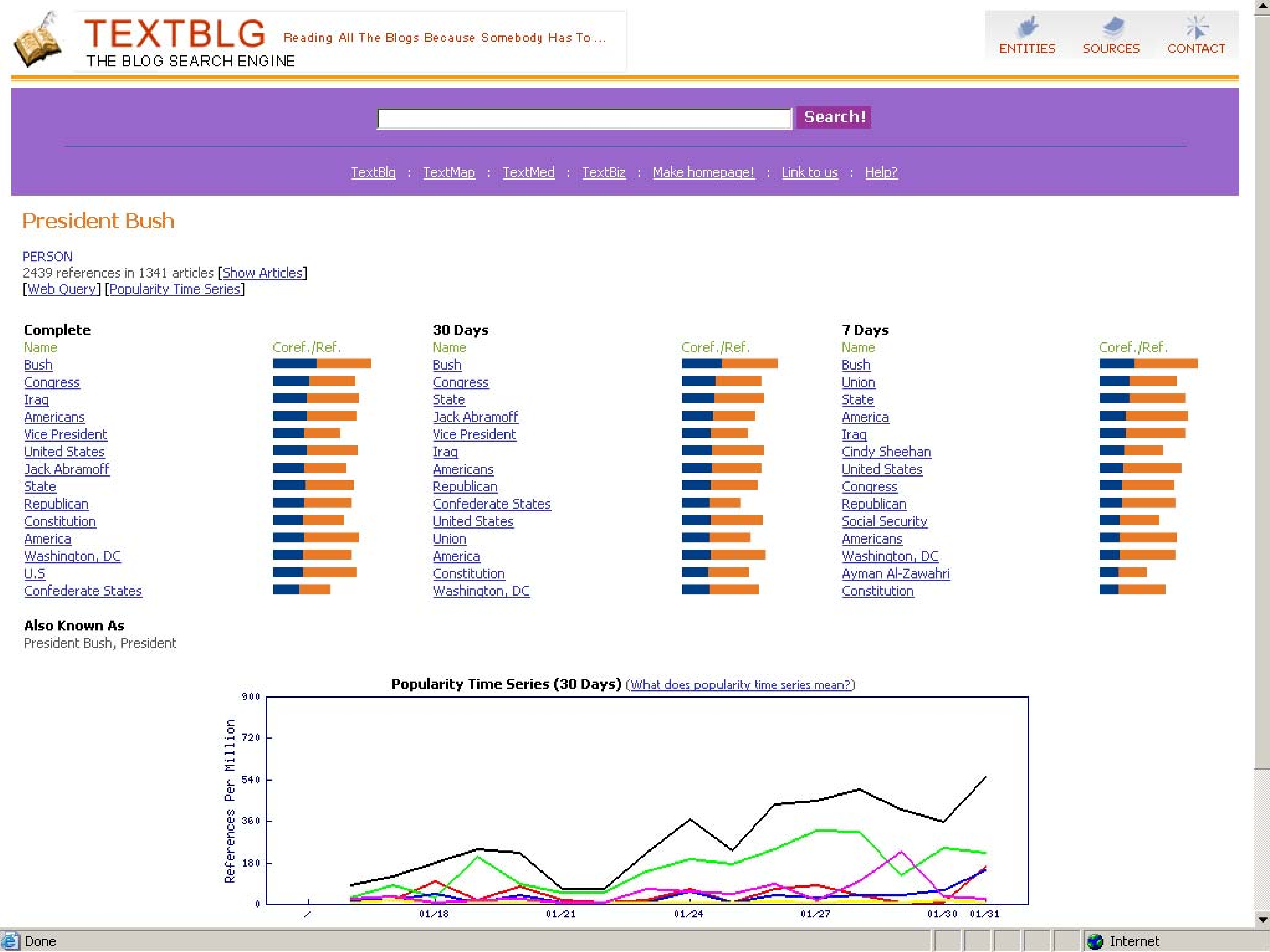
What's New

May 5, 2005
TextMap system goes live!

more news

About TextBlg : TextBlg Team : Disclaimer : Contact Us

Copyright (c) 2005 The Research Foundation of State University of New York
Computer Science Department at Stony Brook University



TEXTBLG Reading All The Blogs Because Somebody Has To ...

THE BLOG SEARCH ENGINE

TextBlg : TextMap : TextMed : TextBiz : Make homepage! : Link to us : Help?

President Bush

PERSON
2439 references in 1341 articles [Show Articles]
[Web Query] [Popularity Time Series]

Complete		30 Days		7 Days	
Name	Coref./Ref.	Name	Coref./Ref.	Name	Coref./Ref.
Bush		Bush		Bush	
Congress		Congress		Union	
Iraq		State		State	
Americans		Jack Abramoff		America	
Vice President		Vice President		Iraq	
United States		Iraq		Cindy Sheehan	
Jack Abramoff		Americans		United States	
State		Republican		Congress	
Republican		Confederate States		Republican	
Constitution		United States		Social Security	
America		Union		Americans	
Washington, DC		America		Washington, DC	
U.S		Constitution		Ayman Al-Zawahri	
Confederate States		Washington, DC		Constitution	

Also Known As
President Bush, President

Popularity Time Series (30 Days) (What does popularity time series mean?)

Outline of Talk

- Lydia NLP pipeline
- Spatial and temporal analysis
- Blogs vs. news
- Current research
- Future visions

System Architecture

Spidering – text is retrieved from a given site on a daily basis using semi-custom spidering agents.

Normalization – clean text is extracted with semi-custom parsers and formatted for our pipeline

Text Markup – annotates parts of the source text for storage and analysis.

Back Office Operations – we aggregate entity frequency and relational data for a variety of statistical analyses.

Levon Lloyd, Dimitrios Kechagias, and Steven Skiena. Lydia: A System for Large-Scale News Analysis. In *String Processing and Information Retrieval: 12th International Conference (SPIRE 2005).*

Text Markup

- We apply natural language processing (NLP) techniques to annotate interesting features of the document.
- Full parsing techniques are too slow to keep up with our volume of text, so we employ shallow parsing instead.
- We can currently markup approximately 2000 newspapers per day per CPU.
- Analysis phases include…

Input

Dr. Judith Rodin, the former president of the University of Pennsylvania, will become president of the Rockefeller Foundation next year, the foundation announced yesterday in New York.

She will take over in March 2005, succeeding Gordon Conway, the foundation's first non-American president. Mr. Conway announced last year that he would retire at 66 in December and return to Britain, where his children and grandchildren live.

Sentence and Paragraph Identification

<p>

Dr. Judith Rodin, the former president of the University of Pennsylvania, will become president of the Rockefeller Foundation next year, the foundation announced yesterday in New York.

</p>

<p>

She will take over in March 2005, succeeding Gordon Conway, the foundation's first non-American president.

Mr. Conway announced last year that he would retire at 66 in December and return to Britain, where his children and grandchildren live.

</p>

Part Of Speech Tagging

<p>

Dr./NNP Judith/NNP Rodin/NNP ,/, the/DT former/JJ president/NN of/IN the/DT University/NNP of/IN Pennsylvania/NNP ,/, will/MD become/VB president/NN of/IN the/DT Rockefeller/NNP Foundation/NN next/JJ year/NN ,/, the/DT foundation/NN announced/VBD yesterday/RB in/IN New/NNP York/NNP./.

</p>

<p>

She/PRP will/MD take/VB over/IN in/IN March/NNP 2005/CD ,/, succeeding/VBG Gordon/NNP Conway/NNP ,/, the/DT foundation/NN 's/POS first/JJ non-American/JJ president/NN ./.

Mr./NNP Conway/NNP announced/VBD last/JJ year/NN that/IN he/PRP would/MD retire/VB at/IN 66/CD in/IN December/NNP and/CC return/NN to/TO Britain/NNP ,/, where/WRB his/PRP$ children/NNS and/CC grandchildren/NNS live/VBP ./.

</p>

Proper Noun Extraction

```
<p>
<pn> Dr./NNP Judith/NNP Rodin/NNP </pn> ,/, the/DT former/JJ president/NN
of/IN the/DT <pn> University/NNP </pn> of/IN <pn> Pennsylvania/NNP </pn> ,/,
will/MD become/VB president/NN of/IN the/DT <pn> Rockefeller/NNP </pn>
Foundation/NN next/JJ year/NN ,/, the/DT foundation/NN announced/VBD
yesterday/RB in/IN <pn> New/NNP York/NNP </pn> ./.
</p>
<p>
She/PRP will/MD take/VB over/IN in/IN March/NNP 2005/CD ,/, succeeding/VBG
<pn> Gordon/NNP Conway/NNP </pn> ,/, the/DT foundation/NN 's/POS first/JJ
non-American/JJ president/NN ./.
<pn> Mr./NNP Conway/NNP </pn> announced/VBD last/JJ year/NN that/IN
he/PRP would/MD retire/VB at/IN 66/CD in/IN December/NNP and/CC return/NN
to/TO <pn> Britain/NNP </pn> ,/, where/WRB his/PRP$ children/NNS and/CC
grandchildren/NNS live/VBP ./.
</p>
```

Actor Classification

<p>

<pn category = "PERSON"> Dr./NNP Judith/NNP Rodin/NNP </pn> ,/, the/DT former/JJ president/NN of/IN the/DT <pn category = "UNKNOWN"> University/NNP </pn> of/IN <pn category = "STATE"> Pennsylvania/NNP </pn> ,/, will/MD become/VB president/NN of/IN the/DT <pn category = "UNKNOWN"> Rockefeller/NNP </pn> Foundation/NN next/JJ year/NN ,/, the/DT foundation/NN announced/VBD yesterday/RB in/IN <pn category = “CITY”> New/NNP York/NNP </pn> ./.

</p>

<p>

She/PRP will/MD take/VB over/IN in/IN <embedded_date> March/NNP 2005/CD </embedded_date> ,/, succeeding/VBG <pn category = "PERSON"> Gordon/NNP Conway/NNP </pn> ,/, the/DT foundation/NN 's/POS <num type = "ORDINAL"> first/JJ </num> non-American/JJ president/NN ./.

<pn category = "PERSON"> Mr./NNP Conway/NNP </pn> announced/VBD last/JJ year/NN that/IN he/PRP would/MD retire/VB at/IN <num type = "CARDINAL"> 66/CD </num> in/IN <embedded_date> December/NNP </embedded_date> and/CC return/NN to/TO <pn category = "COUNTRY"> Britain/NNP </pn> ,/, where/WRB his/PRP$ children/NNS and/CC grandchildren/NNS live/VBP ./.

</p>

Rewrite Rules

<p>
<appellation> Dr. </appellation> <pn category = "PERSON"> Judith Rodin </pn> , the former president of the <pn category = "UNIVERSITY"> University of Pennsylvania </pn> , will become president of the <pn category = "UNKNOWN"> Rockefeller Foundation </pn> next year , the foundation announced yesterday in <pn category = “CITY”> New York </pn> .
</p>
<p>
She will take over in <embedded_date> March 2005 </embedded_date> , succeeding <pn category = "PERSON"> Gordon Conway </pn> , the foundation 's <num type = "ORDINAL"> first </num> non-American president .
<appellation> Mr. </appellation> <pn category = "PERSON"> Conway </pn> announced last year that he would retire at <num type = "CARDINAL"> 66 </num> in <embedded_date> December </embedded_date> and return to <pn category = "COUNTRY"> Britain </pn> , where his children and grandchildren live .
</p>

Alias Expansion

<p>

<appellation> Dr. </appellation> <pn category = "PERSON"> Judith Rodin </pn> , the former president of the <pn category = "UNIVERSITY"> University of Pennsylvania </pn> , will become president of the <pn category = "UNKNOWN"> Rockefeller Foundation </pn> next year , the foundation announced yesterday in <pn category = “CITY”> New York </pn> .

</p>

<p>

She will take over in <embedded_date> March 2005 </embedded_date> , succeeding <pn category = "PERSON"> Gordon Conway </pn> , the foundation 's <num type = "ORDINAL"> first </num> non-American president .

<appellation> Mr. </appellation> <pn category = "PERSON"> Gordon Conway </pn> announced last year that he would retire at <num type = "CARDINAL"> 66 </num> in <embedded_date> December </embedded_date> and return to <pn category = "COUNTRY"> Britain </pn> , where his children and grandchildren live.

</p>

Geography Normalization

<p>

<appellation> Dr. </appellation> <pn category = "PERSON"> Judith Rodin </pn> , the former president of the <pn category = "UNIVERSITY"> University of Pennsylvania </pn> , will become president of the <pn category = "UNKNOWN"> Rockefeller Foundation </pn> next year , the foundation announced yesterday in <pn category = “CITY, STATE, COUNTRY”> New York City, New York, USA </pn> .

</p>

<p>

She will take over in <embedded_date> March 2005 </embedded_date> , succeeding <pn category = "PERSON"> Gordon Conway </pn> , the foundation 's <num type = "ORDINAL"> first </num> non-American president .

<appellation> Mr. </appellation> <pn category = "PERSON"> Gordon Conway </pn> announced last year that he would retire at <num type = "CARDINAL"> 66 </num> in <embedded_date> December </embedded_date> and return to <pn category = "COUNTRY"> Britain </pn> , where his children and grandchildren live.

</p>

Back Office Operations

- The most interesting analysis occurs after markup, using our MySQL database of all occurrences of interesting entities.
- Each day’s worth of analysis yields about 10 million occurrences of about 1 million different entities, so efficiency matters...
- Linkage of each occurrence to source and time facilitates a variety of interesting analysis.

Duplicate Article Elimination

Supreme Court Justice David Souter suffered minor injuries when a group of young men assaulted him as he jogged on a city street, a court spokeswoman and Metropolitan Police said Saturday.

Supreme Court Justice David Souter suffered minor injuries when a group of young men assaulted him as he jogged on a city street, a court spokeswoman and Metropolitan Police said.

Hashing techniques can efficiently identify duplicate and near-duplicate articles appearing in different news sources.

Synonym Sets

- JFK, John Kennedy, John F. Kennedy, and John Fitzgerald Kennedy all refer to the same person.
- We need a mechanism to link multiple entities that have slightly different names but refer to the same thing.
- We say that two actors belong in the same synonym set if:
  - There names are morphologically compatible.
  - If the sets of entities that they are related to are similar.

Levon Lloyd, Andrew Mehler, and Steven Skiena. Identifying Co-referential Names Across Large Corpra. In *Proc. Combinatorial Pattern Matching (CPM 2006)*

Outline of Talk

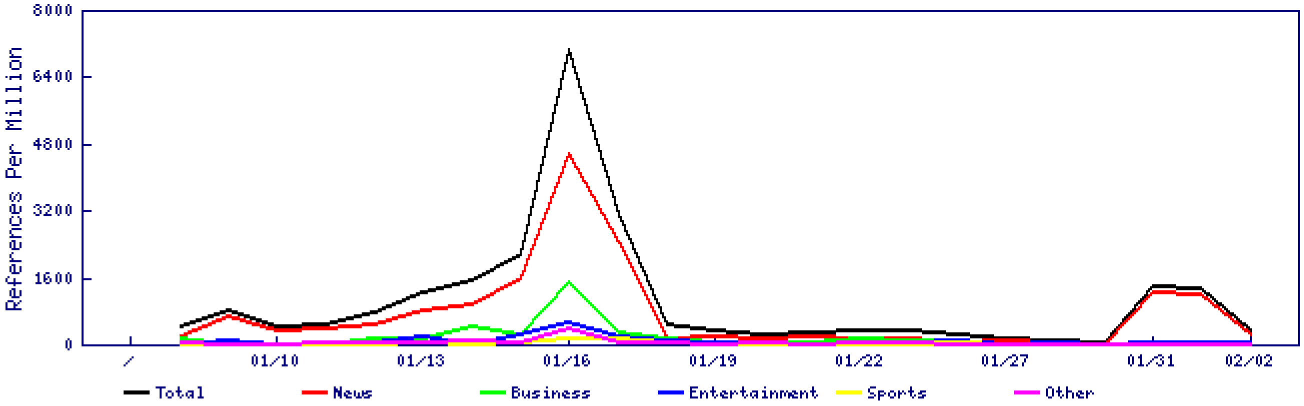
- Lydia NLP pipeline
- Spatial and temporal analysis
- Blogs vs. news
- Current research
- Future visions

Juxtaposition Analysis

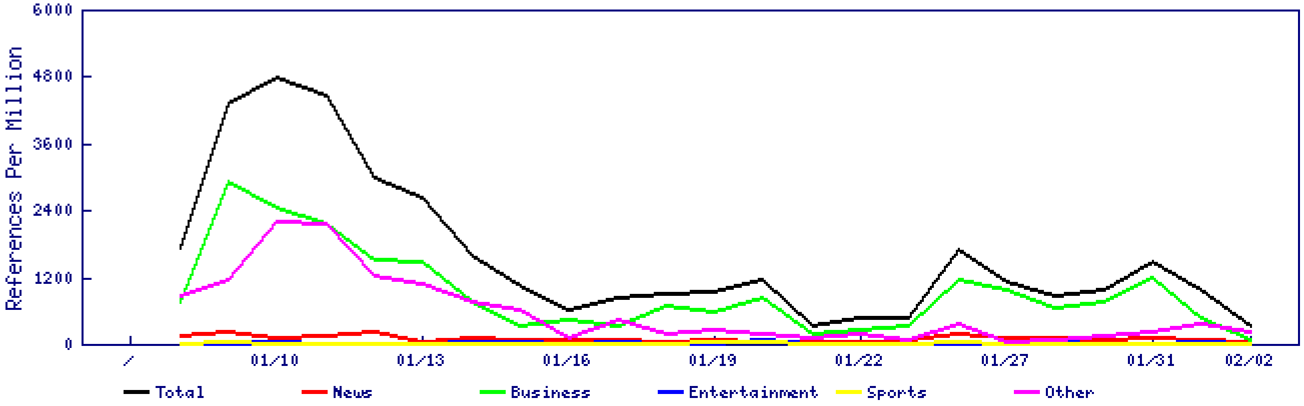
- We want to compute the significance of the co-occurrences between two entities
- Similar to *collaborative filtering, determining which customers are most similar in order to predict future buying preferences*
- Just counting the number of co-occurrences causes the most popular entities to be related to everyone

Time Series Analysis

Martin Luther King



Samuel Alito

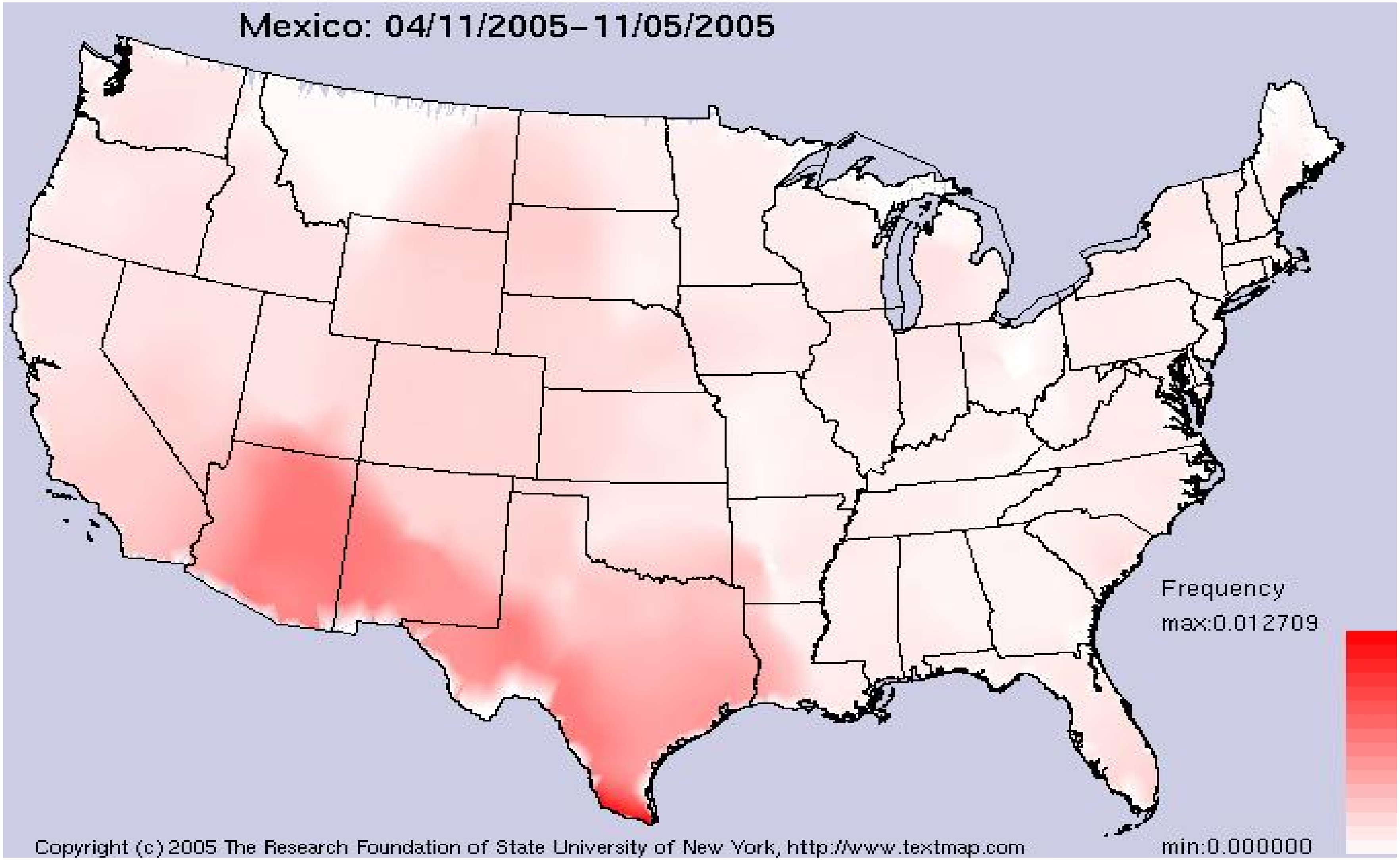


Heatmaps

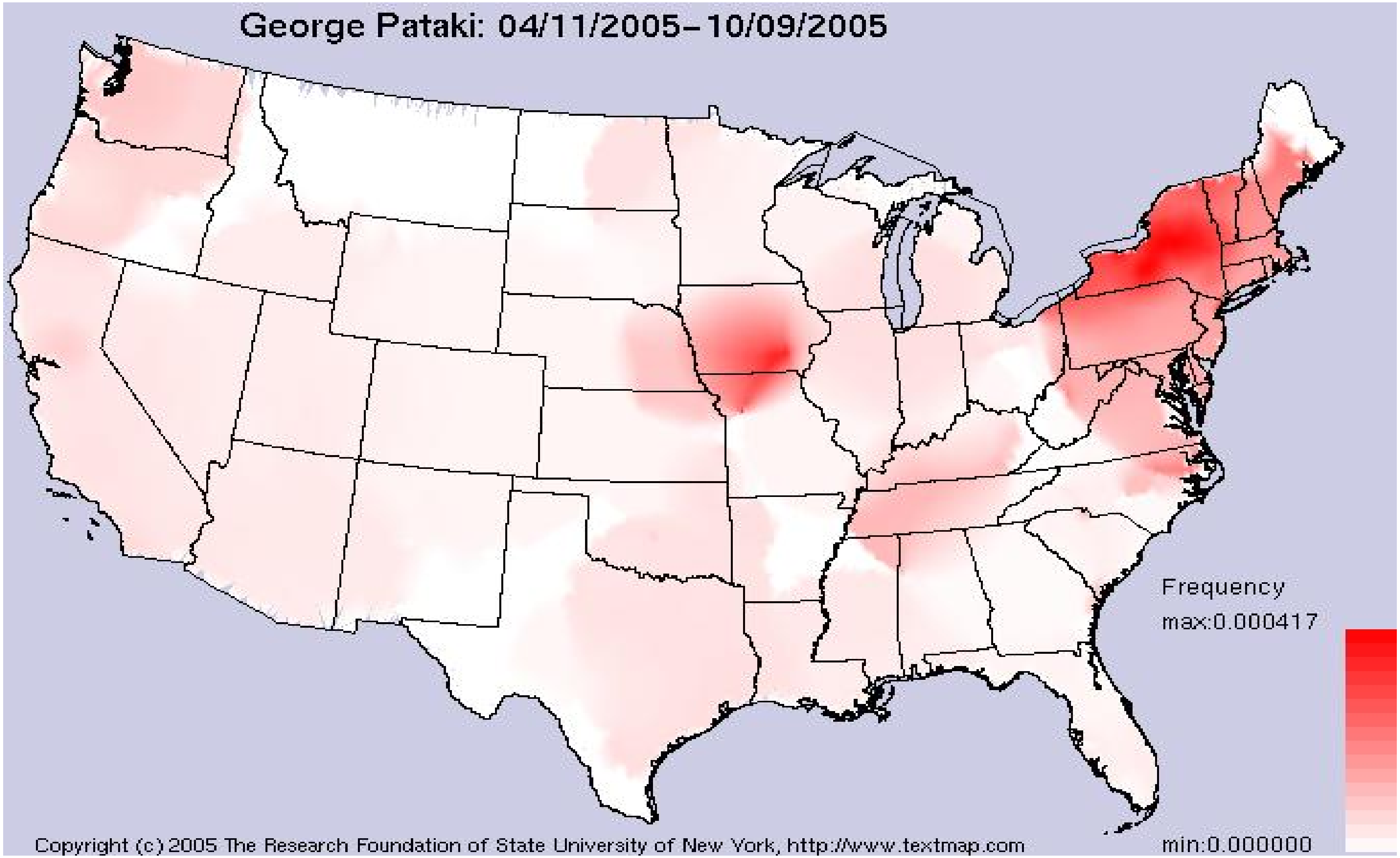
- Where are people are talking about particular topics?
- Newspapers have a *sphere of influence based on:*
  - *Power of the source – circulation, website popularity*
  - *Population density of surrounding cities*
- The *heat* a given entity generates in a particular location is a function of the frequency it is mentioned in local sources

A. Mehler, Y. Bao, X. Li, Y. Wang, and S. Skiena. Spatial analysis of News Sources, *IEEE Trans. Visualization* (2006)

Donde Esta Mexico?


Mexico: 04/11/2005–11/05/2005
Frequency
max:0.012709
min:0.000000
Copyright (c) 2005 The Research Foundation of State University of New York, http://www.textmap.com

Who is running for president?


George Pataki: 04/11/2005–10/09/2005
Frequency
max:0.000417
min:0.000000
Copyright (c) 2005 The Research Foundation of State University of New York, http://www.textmap.com

New Orleans – Animation

TEXTMAP

THE ENTITY SEARCH ENGINE

Monitoring the World So You Don't Have To ...

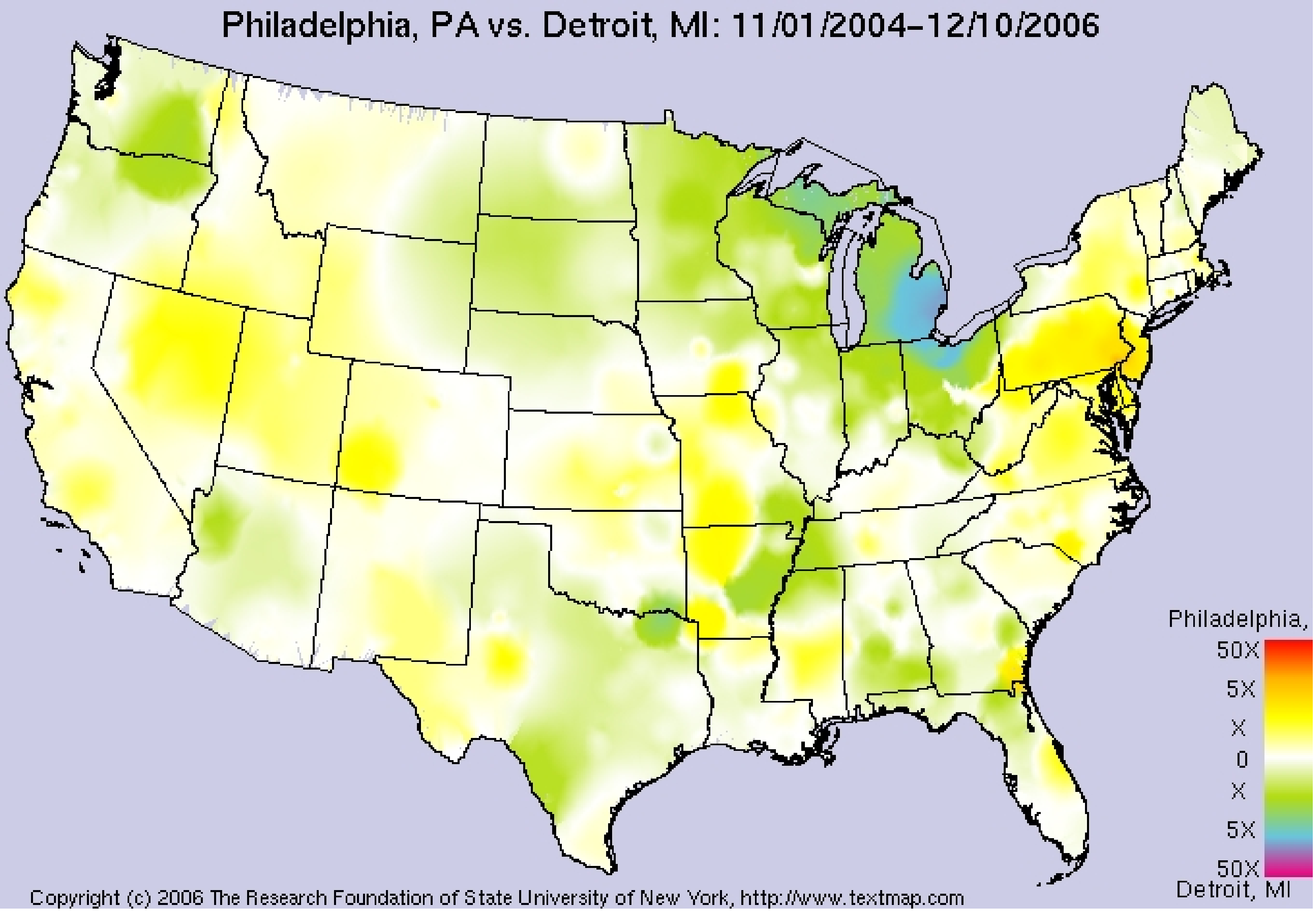
Monthly Spatial News Analysis

New Orleans, LA

11/01/2004 - 12/01/2006

http://www.textmap.com

Comparative Entity Maps


Philadelphia, PA vs. Detroit, MI: 11/01/2004–12/10/2006
Philadelphia,
50X
5X
X
0
X
5X
50X
Detroit, MI
Copyright (c) 2006 The Research Foundation of State University of New York, http://www.textmap.com

Outline of Talk

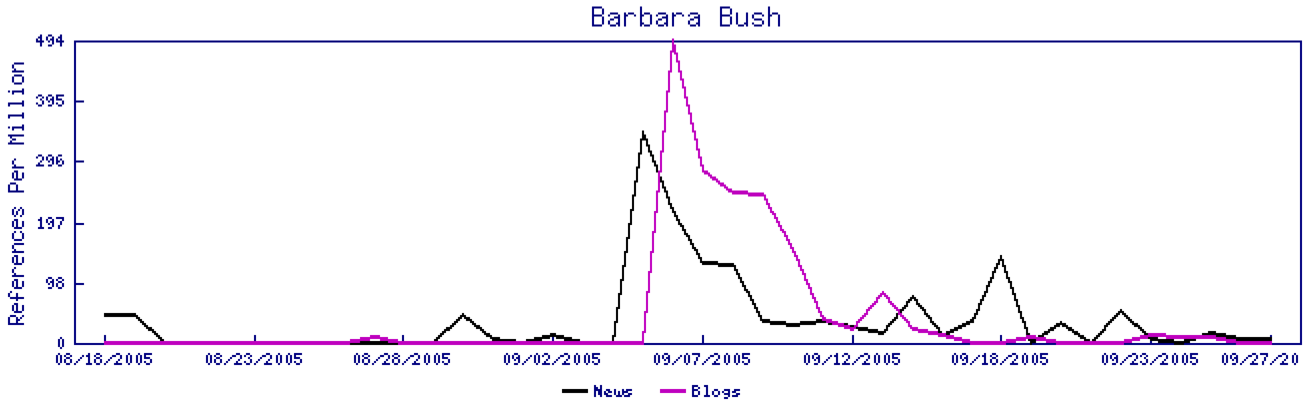
- Lydia NLP pipeline
- Spatial and temporal analysis
- Blogs vs. news
- Current research
- Future visions

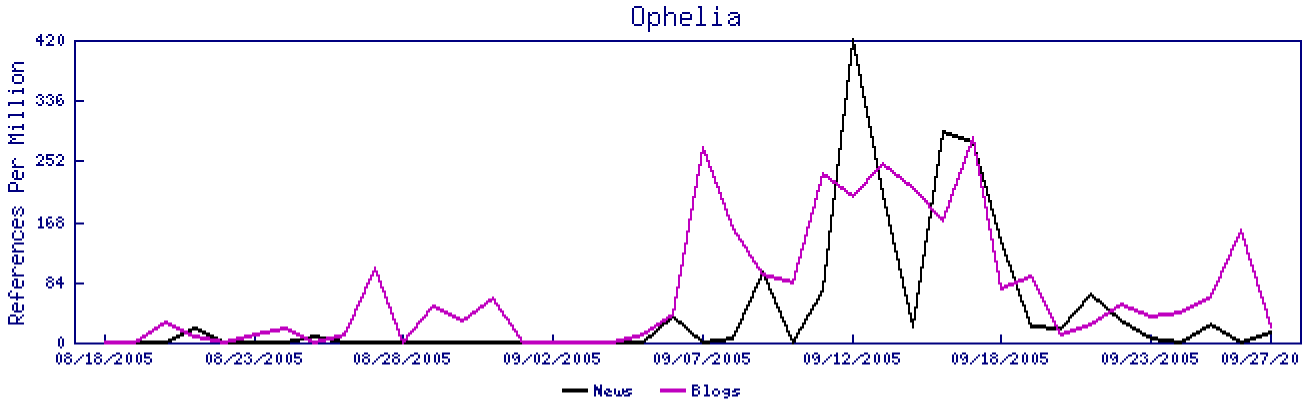
Blog Analysis with Lydia

- Blogs represent a different view of the world than newspapers.
  - Less objective
  - Greater diversity of topics
- We adapted Lydia to process *Livejournal* blogs, and compared blog content to that of newspapers.

Levon Lloyd, Prachi Kaulgud, and Steven Skiena. News vs. Blogs: Who Gets the Scoop?.
In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs.*

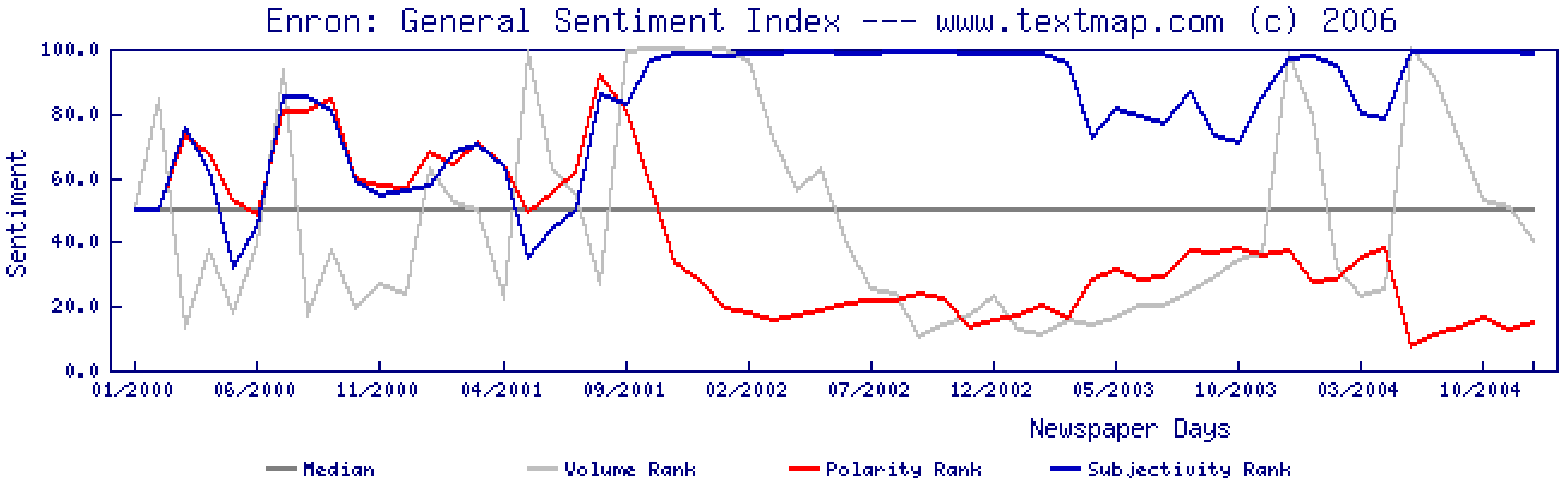
Who Gets the Scoop?





Sentiment Analysis

- Sentiment analysis lets us to measure how positively/negatively an entity is regarded, not just how much it is talked about.



Most Positive Actors in News and Blogs

- **News:** Felicity Huffman, Fenando Alonso, Dan Rather, Warren Buffett, Joe Paterno, Ray Charles, Bill Frist, Ben Wallace, John Negroponte, George Clooney, Alicia Keys, Roy Moore, Jay Leno, Roger Federer
- **Blogs:** Joe Paterno, Phil Mickelson, Tom Brokow, Sasha Cohen, Ted Stevens, Rafael Nadal, Felicity Huffman, Warren Buffett, Fernando Alonso, Chauncey Billups, Maria Sharapova, Earl Woods, Kasey Kahne, Tom Brady

Most Negative Actors in News and Blogs

- **News**: Slobodan Milosevic, John Ashcroft, Zacarias Moussaoui, John Allen Muhammad, Lionel Tate, Charles Taylor, George Ryan, Al Sharpton, Peter Jennings, Saddam Hussein, Jose Padilla, Abdul Rahman, Adolf Hitler, Harriet Miers, King Gyanendra
- **Blogs:** John Allen Muhammad, Sammy Sosa, George Ryan, Lionel Tate, Esteban Loaiza, Slobodan Milosevic, Charles Schumer, Scott Peterson, Zacarias Moussaoui, William Jefferson, King Gyanemdra, Ricky Williams, Ernie Fletcher, Edward Kennedy, John Gotti

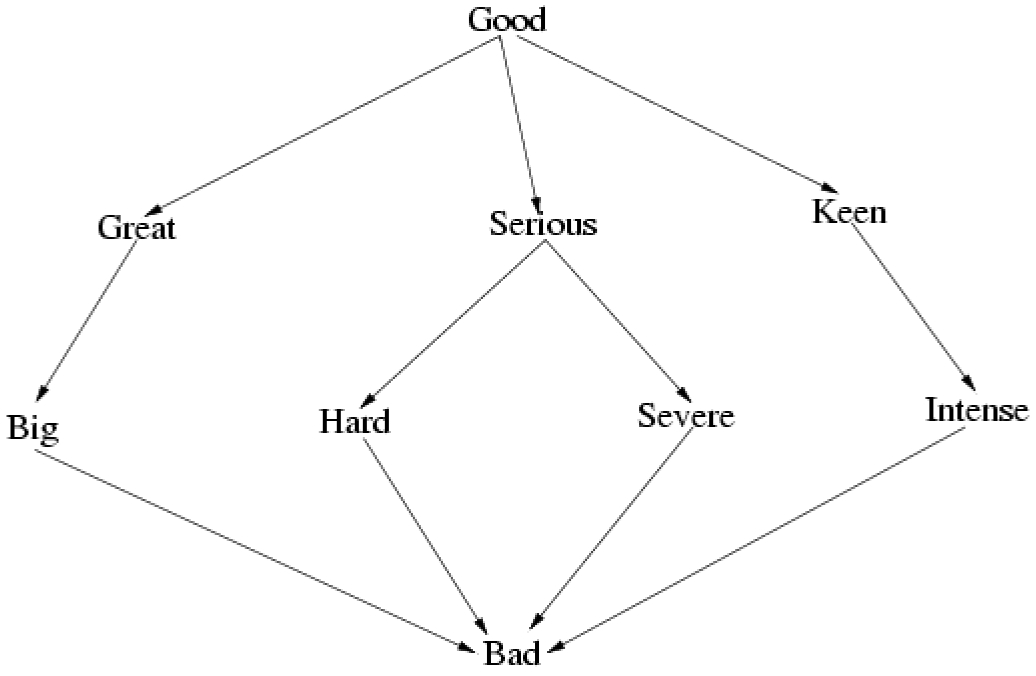
How Do We Do it?

- We use large-scale statistical analysis instead of careful NLP of individual reviews.
- We expand small seed lists of +/- terms into large vocabularies using Wordnet and path-counting algorithms.
- We correct for modifiers and negation.
- Statistical methods turn these counts into indicies.

N. Godbole, M. Srinivasaiah, and S. Skiena. Large-Scale Sentiment Analysis for News and Blogs. *Int. Conf. Weblogs and Social Media*, 2007

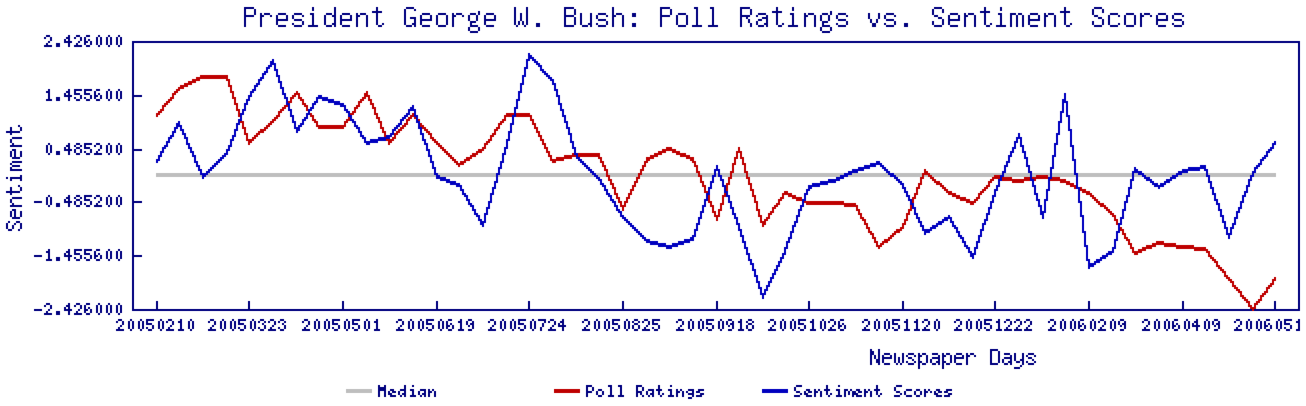
Good to Bad in Three Hops

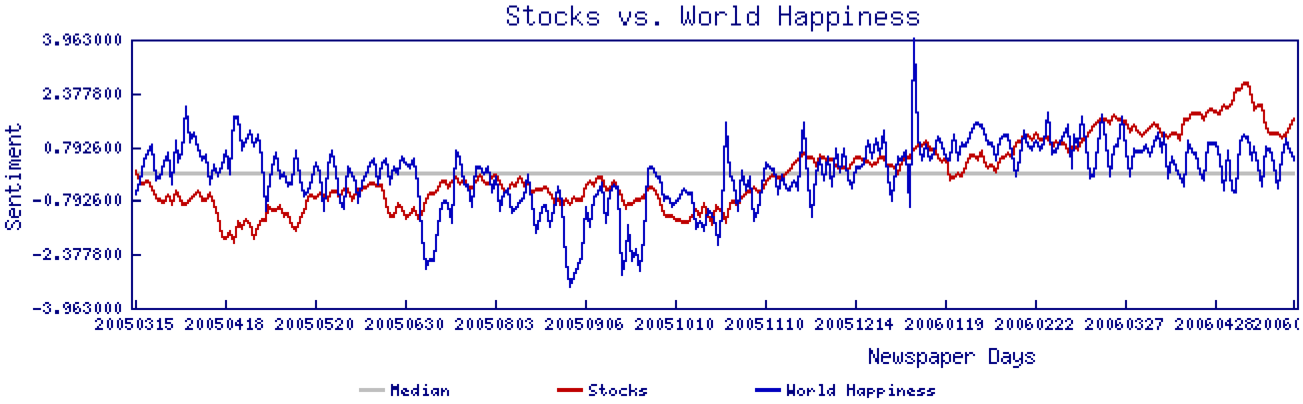
- Paths of WordNet synonyms can lead to contradictory results, requiring careful path selection.



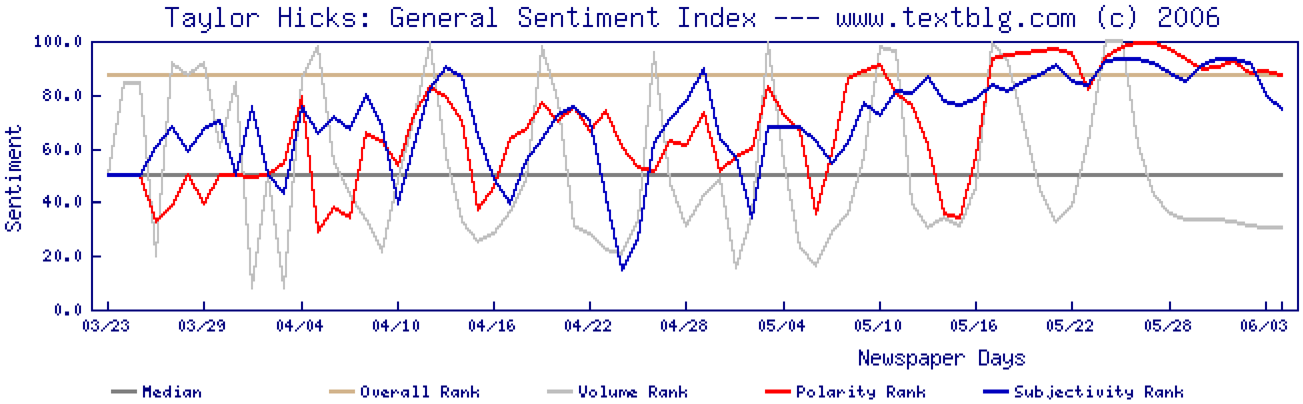
What Does it Mean?

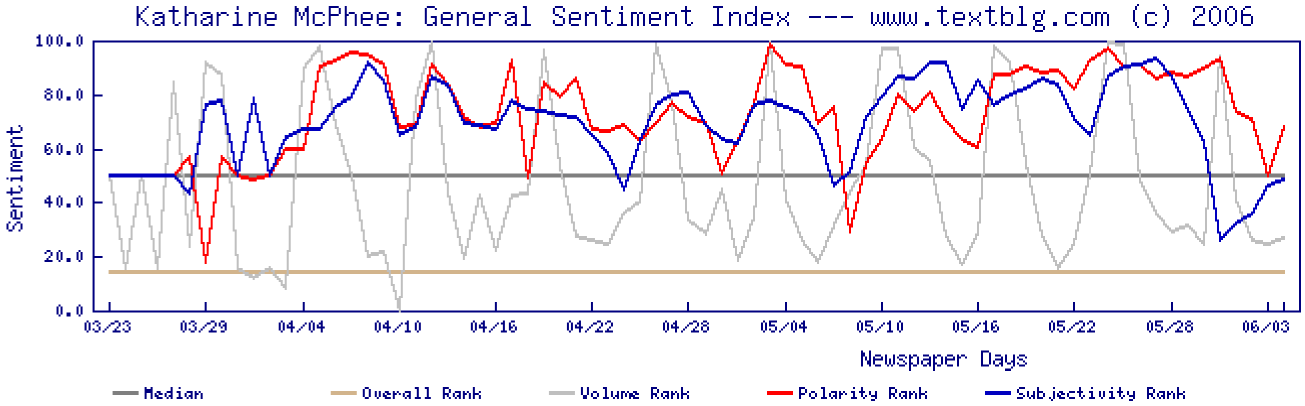
- Our scores corrolate very well with financial, political, and sporting events.





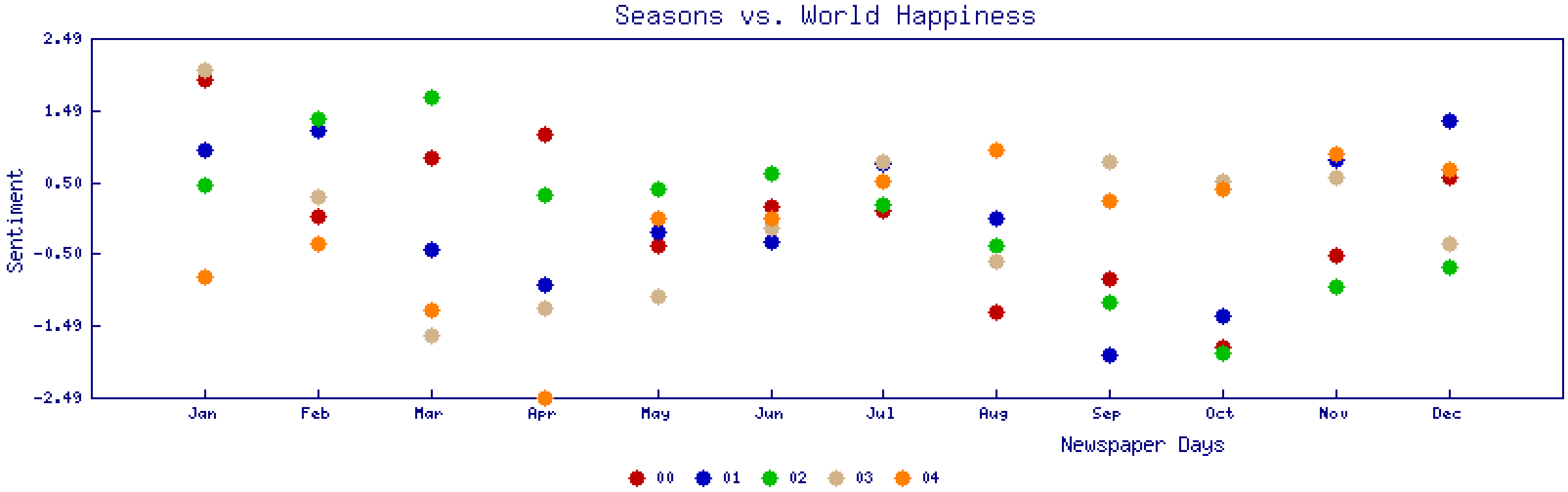
Who is the *American Idol*?



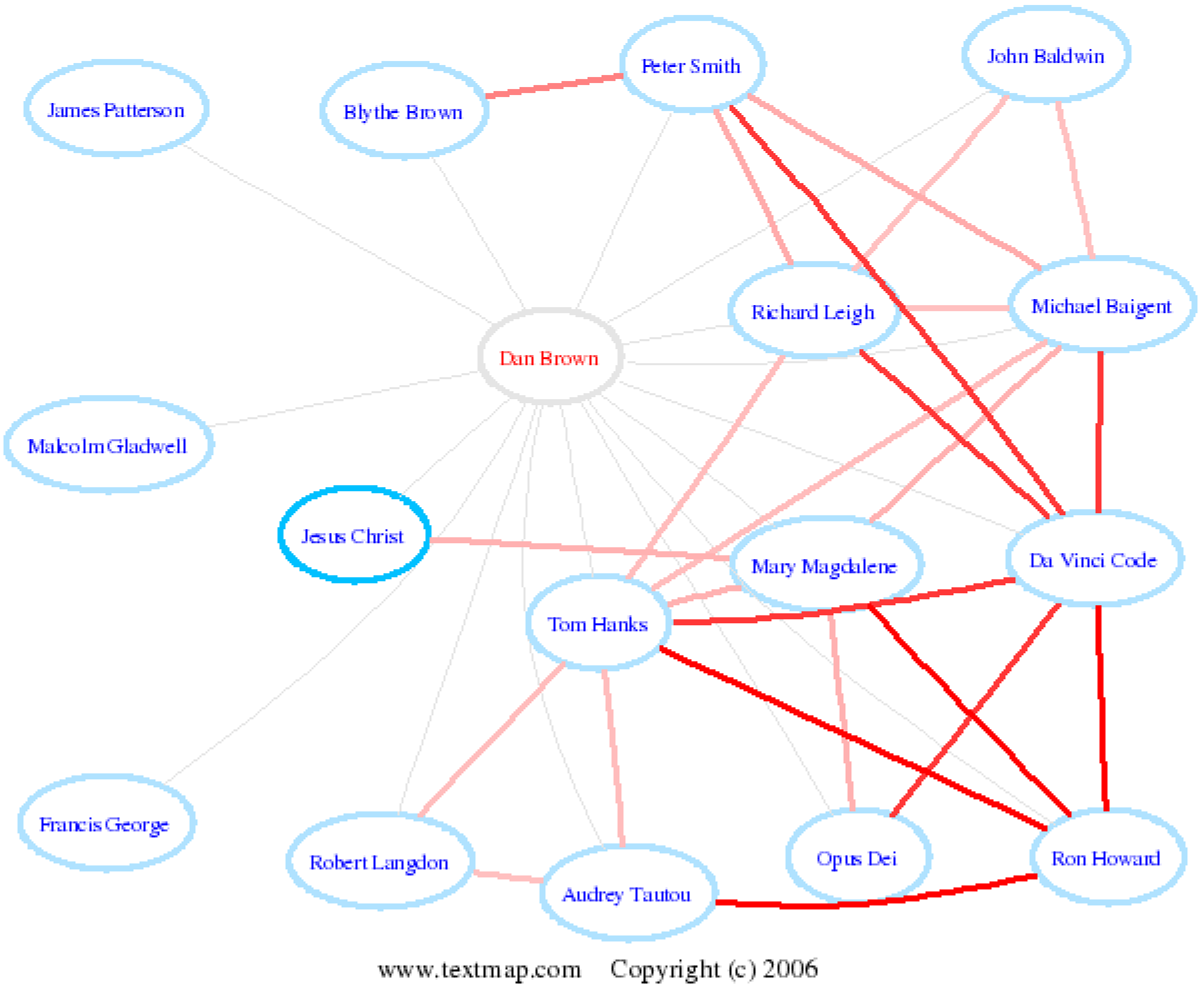


Seasonal Effects on Sentiment

- The low point is not September 2001 but April 2004, with the Madrid bombings and war in Iraq.



Social Network Analysis



Relationship Identification

- We use verb-frames and template-based methods to try to identify the nature of statistically-significant relationships, e.g
- devastated <Hurricane Katrina:Louisiana>
- killed-in <Diana:Paris, FRA>
- became <Joseph Ratzinger:Pope Benedict XVI>
- not-watch <Dalai Lama:`` The Simpsons ">

Description Extraction

- We use template-based methods and WordNet sense analysis to extract meaningful descriptions, such as:
- Warren Buffett, billionaire investor
- Giacomo, Kentucky Derby winner
- Kim Jong Il, North Korean leader

Outline of Talk

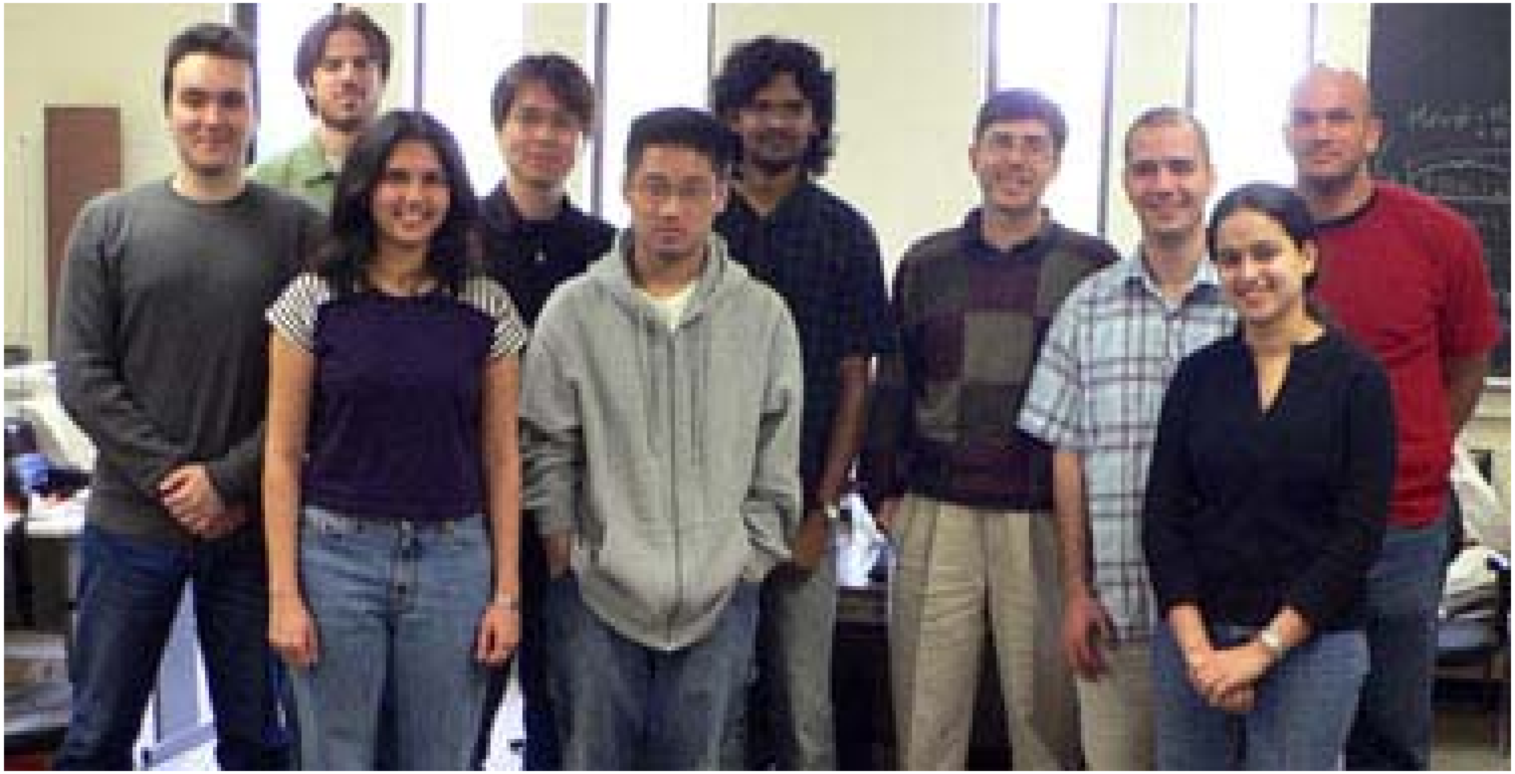
- Lydia NLP pipeline
- Spatial and temporal analysis
- Blogs vs. news
- Current research
- Future visions

Future Directions

- Entity-oriented (instead of document-based) search engines
- Foreign-language news analysis
- Event-focused relation extraction
- Financial modelling and analysis
- Social network analysis

We actively seek collaboration with social scientists

The Lydia Team



… and the Lydia Cluster