

# Declaring Independence from the President in the 2006 Mid-term Elections<sup>1</sup>

Prepared for delivery at the 2007 Annual Meeting of the Midwest Political Science Association

Chicago, IL. April 14-17.

Stephen Purpura  
The Program on Networked Governance  
John F. Kennedy School of Government - Harvard University  
[stephen\\_purpura@ksg07.harvard.edu](mailto:stephen_purpura@ksg07.harvard.edu)

David Lazer  
The Program on Networked Governance  
John F. Kennedy School of Government - Harvard University  
[David\\_Lazer@Harvard.edu](mailto:David_Lazer@Harvard.edu)

Kevin Esterling  
University of California Riverside  
[kevine@citrus.ucr.edu](mailto:kevine@citrus.ucr.edu)

Dustin Hillard  
Program on Networked Governance – Harvard University  
Dept of Electrical Engineering - University of Washington  
[hillard@u.washington.edu](mailto:hillard@u.washington.edu)

Michael Neblo  
Ohio State University  
[Neblo.1@polisci.osu.edu](mailto:Neblo.1@polisci.osu.edu)

---

<sup>1</sup> We acknowledge support from the Program on Networked Governance at Harvard through NSF grant No. 0429452, the Connecting to Congress Project, and the Congressional Management Foundation. The views expressed are those of the authors and not the National Science Foundation. We would also like to thank our dedicated research assistants, Matt Vandenberg, Darren Mitchell, Ray Amberg, Andre Clark, and Frida for their outstanding work in support of this paper and Eduard Hovy and David King for their helpful comments and feedback.

## **Abstract**

Did Republican Members of Congress "run away" from President Bush? We examine mentions of President Bush on official Member websites to evaluate Members' communication strategies. This analysis builds on a methodological innovation to identify mentions of Bush from two terabytes of snapshots of Member websites from 2001 to 2006. A combination of natural language processing techniques, including structured learning, active learning and sentiment analysis, facilitate data generation. For the first time in political science, the use of NLP sentiment analysis with temporal induction (time based decay for topic drift) also appears.

Within this early work-in-progress, we assess positive and negative mentions of Bush, assess the importance of electoral vulnerability while controlling for other characteristics of the Member and his/her district. Using these new methods, we believe we will be able to relate the signals measured in the data generation process to the dynamics of campaign strategies.

## Introduction

In 1998, few Congressional Members had web sites, let alone official web sites which they updated daily. By 2002, an estimated 70% of Congressional candidates had web sites (Jacobson, 2004; p: 89). In 2006, most Congressional Members have official web sites and they update them frequently. Some of these updates are recorded at the Internet Archive ([www.archive.org](http://www.archive.org)), which attempts to take frequent snapshots of how web sites appeared at a specific point in time. The Archive has snapshots of many web sites, including a November 11, 1998 prototype of Google that you can visit.

The existence of official websites creates a systematic record of Congressional communications. This provides a goldmine for researchers interested in the strategic communication decisions of Members. What are Members, collectively, talking about? How do their messages evolve with events?

Examination of these data, however, poses significant practical challenges. In particular, the sheer volume overwhelms traditional human coding, where, for example, the September 2001 through December 2006 snapshots of the Congressional web sites stored at [www.house.gov](http://www.house.gov) and [www.senate.gov](http://www.senate.gov) require more than 2 *terabytes* of storage space. Further, we would note, the incremental volume is increasing enormously, because audio and video (memory intensive data) is increasingly being placed on the website. It would potentially take multiple lifetimes for human coders to sift through this pile.

This paper seeks to surmount this methodological challenge by demonstrating the effectiveness of information extraction from text using a combination of computer software, including natural language processing technology (NLP), integrated into a context analysis process. Further, this paper provides the first application of *NLP sentiment analysis with temporal induction* in political science. This new technology is related to industry research work on managing topic or concept drift during topic classification of time series data sets (Forman, 2006) and its goal is to reduce the number of times that computer software mistakenly relates similar words in different topics for each other. An example would be the incorrect classification of “choice” (as it relates to abortion) for “choice” (as it relates to voting). In our work, this distinction is important when ‘choice’ (abortion) is linked negatively with the President on some web sites during 2003 and ‘choice’ (voting) is linked positively with the President on other web sites during 2006. Although this system certainly doesn’t solve all of the problems of concept drift, evidence suggests it solves some problems and improves accuracy, so we deploy it here as a part of testing it for inclusion in the bulk of our work.

Substantively, we use this method to examine how Members communicate their views regarding the incumbent President, and how that characterization is affected by (1) party affiliation; (2) popularity of incumbent President; and (3) proximity of election. In the particular application we use automated coding to evaluate whether a Member posts positive, negative, or neutral messages about President George W. Bush. We find that Members are strategic in their depiction of the President, where overall criticism of the President tracks closely with Presidential popularity, where, in particular, there is a notable split in behavior among Republicans as the 2006 election neared: invulnerable

Republicans continued to say positive things about the President, vulnerable Republicans abruptly fled the President, attempting to differentiate themselves from President Bush. These results both highlight a particular pattern of strategic communication by Members as well as the potential of this data set combined with NLP to produce insights about the communication strategies of Members of Congress.

### **Strategic Communication by Members of Congress**

It's difficult to discuss elections without discussing uncertainty, and Congressional Members seem to know this (Jacobson, 2004; p. 87). Members have incentives to use strategic communication methods that work, but they are uncertain which methods work and don't work (Jacobson, 2004; p. 87). Despite this uncertainty, they invest time and money in a variety of activities and they take advantage of incumbent advantages like franking. Franked Congressional mail doesn't usually contain stories about the policy failures the Member supported with quips like, "We'll get it right next time!" So we assume a limited form of rationality such that most Members control staff and seek to place primarily strategic messages on their official web sites – even if that message is as simple as, "I'm available to help you!"

In keeping with Jacobson's theory about campaign messages, campaigns seek to develop a consistent theme which frames the choice in a positive light for the candidate (91). The primary reason to believe the theme wouldn't follow through to the Internet is a lack of faith in constituent use. Doubt may have ruled in 1998, but by the 2002 election season (when an estimated 70% had campaign web sites) it is increasingly difficult to accept. The lone holdouts were probably the least vulnerable Members.<sup>2</sup>

Moreover, since challengers frequently seek to hang incumbents with failed policies and their association with unpopular parties, people, and events (Jacobson, 2004; p. 92), it would seem ideal to have a place to list accomplishments and focus on positive constituent messages which counter these attacks. Similarly, since incumbents seek to "remain above the fray" as long as possible (Jacobson, 2004: p. 93), it is dubious to assume that they will "go negative" in their official communication without good reason. In the case of our investigation, leveling attacks at popular people seems like poor strategy.

One of the key mechanisms that Members of Congress have in "branding" themselves is in defining where they stand vis a vis the President. Members, even longstanding incumbents, face the odd relationship with the President that the President is far more well-known to their constituents than they are. Well-known, of course, does not mean well liked. One efficient mechanism, therefore, for a Member to convey the type of Representative she or he is, is by conveying whether they stand with or against the President. Of course, some of this is inescapable, due to party affiliation. However, some of

---

<sup>2</sup> We have evidence that the lone hold-outs were the least vulnerable members. We should add this data to a future version of the paper.

this is manipulable through the content of what the Member communicates to constituents through a variety of means. Looking at figure 1, assume that a given constituent has a view of the President (positive or negative) that is given, where the Member can manipulate the perception of constituents of the valence of their connection to the President. The constituent's view, in turn, would be driven by their view of the President and of the connection between the Member and President. A Member who is positively connected to a President that one likes is generally also going to be viewed positively. Conversely, one that is positively associated with a President that one dislikes is also likely to be disliked.

[figure 1 here]

We assert that this positive/negative connection to the President will therefore be strategically manipulated by Members so as to increase the probability that their constituents will view them positively. Further, we would argue that the websites offer an efficient, if indiscriminate, way to reach constituents. We would therefore expect that the depiction of the President will vary with the dominant views of the President among constituents.

*Proposition 1: The valence of the depiction of the President will be strongly associated with the broad popularity of the President among constituents.*

Further, while these strategic calculations are always being made, we assert that it is most important when electoral considerations are pre-eminent, specifically for vulnerable incumbents near an election.

*Proposition 2: The strategic calculus as to how to depict relationship to the President is heightened with the proximity to the election and the vulnerability of the incumbent.*

### **Official Web Sites of Members of Congress**

Each Congressional Member is given an official web site at [www.House.gov](http://www.House.gov) or [www.Senate.gov](http://www.Senate.gov) which is designed to enable communication between a Member and their constituents. This web site enables the Member to not only convey important happenings in Congress but also take feedback and questions from constituents. The Congressional Management Foundation ([www.cmfweb.org](http://www.cmfweb.org)) publishes "The Gold Mouse Report" which ranks the effectiveness of official Congressional web sites at accomplishing these goals.<sup>3</sup>

In contrast to *campaign* web sites, official Congressional web sites are not allowed to solicit campaign contributions or distribute campaign messages. However, like the franking privilege, Members are allowed to use their web real estate to control the agenda of communication with

---

<sup>3</sup> The 2006 Congressional Management Foundation Gold Mouse Report can be downloaded at [http://www.cmfweb.org/storage/cmfweb/documents/CMF\\_Pubs/2006GoldMouseReport.pdf](http://www.cmfweb.org/storage/cmfweb/documents/CMF_Pubs/2006GoldMouseReport.pdf)

constituents who are increasingly using the Internet to search for information about the activities of the government (Pew Internet, 2007)<sup>4</sup>.

In our case, we assume that the home page of the Member has special agenda setting properties not unlike the top left hand column of the front page of the New York Times and Washington Post. This space is the first image that a constituent sees when they visit the Member's site, and most members use the opportunity to highlight timely messages to their constituents (see Gold Mouse Report). In addition, cheap storage costs coupled with high transaction costs in web site management encourage the Member to leave documents on their web site for a long period. Other than enabling opposition research, there is little incentive for a Member to remove linked documents except to control messaging to constituents.

During a campaign season, the Member has many different avenues to contact their constituents. These include public appearances, television commercials, mailings, email newsletters, campaign web sites and official web sites. For the purposes of this work, we just focus on observing the activities at the Member web sites, both prior to the 2006 midterm election and during the election cycle, to document the patterns and then examine them against theories of political behavior.

## **Observing Congressional Member Web Sites**

Gathering data on official Congressional Member sites is a daunting task and there are really two methods of completing the activity. One method asks trained human coders to conduct content analysis of the sites according to a codebook. The other seeks to have a computer conduct the analysis. In our case, we use both methods. We ask trained human coders to conduct analysis with the idea that their effort can be translated into leverage by the computer. Similar systems have been shown to reduce the total amount of content analysis work required by humans by more than 80%, even when starting from scratch (Hillard, Purpura, and Wilkerson, 2007).

For this work, we require more labeling by humans than should be required to facilitate systemic analysis of the method. A 'snapshot' of a Congressional web site is a mostly working version of the exact web site as it appeared on some date. Humans use any web browser to view the list of snapshots and analyze the contents of the home page, searching for mentions of President Bush and recording the sentiment of each mention. Samples of the Congressional web site snapshots are available at [www.connectingtocongress.org](http://www.connectingtocongress.org), a web site which is maintained by the project team.

The snapshots are taken from the Internet Archive ([www.archive.org](http://www.archive.org)) and directly from the Congressional Member web sites themselves using a technique known as 'web crawling'. Our web

---

<sup>4</sup> See Pew Internet's January 2007 polling results from the 2006 election cycle.  
[http://www.pewinternet.org/pdfs/PIP\\_Politics\\_2006.pdf](http://www.pewinternet.org/pdfs/PIP_Politics_2006.pdf)

crawling program stores the snapshots at the Connecting to Congress web site for analysis using other computer programs written for the purpose<sup>5</sup>. As mentioned, we require more human analysis because, unlike Hillard, Purpura, and Wilkerson (2007), we attempt to extract opinions (sentiment) from the text and this operation is considered more complex than extracting information such as the primary topic under discussion (Breck, Choi, and Cardie, 2007; Choi, Breck, and Cardie, 2006; Stoyanov and Cardie, 2006; Stoyanov and Cardie, 2006; Wiebe, Wilson, Cardie, 2005; Choi, Cardie, Riloff, and Patwardhan, 2005). Since sentiment analysis is more complex and it has never been completed on official Congressional web sites before, we assume that we don't know the exact difficulties that we will encounter.

Even for professional computer scientists and engineers this project is ambitious. First, the amount of data involved is very large, in our case terabytes.<sup>6</sup> Second, many web sites are constructed using dynamic technology which make them adaptive to the people that visit them. Finally, copyright issues associated with non-governmental web sites can complicate storage and analysis. In our case, we deal with these issues in a number of ways. We admit that we greatly benefit because several members of our project team are professional software developers<sup>7</sup>, trained in computer science and information technology. This reduces the problem from one of unknown complexity to one of time and money management. Additionally, copyright issues are not significant for web site content at [www.house.gov](http://www.house.gov) and [www.senate.gov](http://www.senate.gov) because these web sites are considered public content which are not subject to copyright enforcement. However, we do not grab data from non-governmental web sites unless we tell the system by hand that the web site is a Congressional member site<sup>8</sup>.

In total, we established storage mechanisms capable of capturing 1 million web site snapshots. The math for this is fairly simple:

$$1 \text{ million} = (435 \text{ Congressional Members} + \text{Committee Sites}) * 365 \text{ days of the year} * 6 \text{ years}$$

Since the Internet Archive does not have snapshots of every Congressional member on every day of the year, we don't need this much storage. However, we frequently captured more than one snapshot in a

---

<sup>5</sup> We would like to thank Darren Mitchell, Dustin Hillard, Stephen Purpura, Ray Amberg, and Andre Clark for their assistance in writing the software associated with this project. The software will be released to the public in separate publications.

<sup>6</sup> According to the SI standard a terabyte is  $10^{12}$  bytes or the equivalent of 1,000 gigabytes. In traditional computer science memory allocation terms, a terabyte is  $2^{40}$  bytes, which is now referred to as a tebibyte. Regardless, it is more data than a typical desktop computer can store at this time.

<sup>7</sup> Two members of our project team worked at Microsoft on the Internet Explorer browser together, which gives them a unique insight into how web pages are rendered and how text can be extracted.

<sup>8</sup> Some Congressional Members store their official sites at URLs which do not match [www.senate.gov](http://www.senate.gov) or [www.house.gov](http://www.house.gov).

day for testing purposes. In total, we stored more than 250,000 snapshots, requiring more than 2 terabytes of data. Once we've captured all of the sites, we had to design computer software to 'reconstruct' them for re-display to the public. The web site at [www.ConnectingtoCongress.org](http://www.ConnectingtoCongress.org) shows the results of this work, which is amazingly capable of allowing users to browse past sites as if they were still live.

In addition to the changes we make for display of the snapshots, we also make duplicate copies of the HTML versions of the web sites using special tools developed by Purpura, Mitchell, and Hillard. These tools dice the HTML into sections, preserving section headings and discriminating sections of text. These text sections represent the text comments on the web sites, as intended by Members to be seen by their constituents. Our processing methods simply format the Members statements into text files that are easy for the computer software to process. Links between the web sites, the URLs, Congressional Member attribute data, and all of the files are maintained in hashes and relational databases.

## **Analyzing Congressional Member Web Sites Using a Computer**

We code web sites for the purpose of generating data for analysis. It is fairly straight-forward to understand how humans analyze the web sites. We have a small codebook for them, they follow it, and we check it. However, for this work we attempt to improve on humans by leveraging computers in the process to save time in future labeling experiments. For this, we take the processed text files and feed them to structured learning systems.

Structured learning systems, which are widely accepted in many fields, have received less attention within Political Science. In Purpura & Hillard (2006), we used a common machine learning technique (support vector machines or SVMs) and cross-validation to demonstrate that a complex topic classification scheme could be automatically applied to congressional bill titles with high fidelity results. Later, in Hillard, Purpura, and Wilkerson (2007), we improve upon this work to turn it into an operational context-analysis tool for researchers to work against complex data sets. In both works, Congressional bill titles – which average only 15 words – enable the system to find enough signals to nearly duplicate the ability of human researchers to sort bills into a complex 226 category system with high levels of reliability. In this work, snippets of Congressional web sites average just slightly more than 18 words and we are now trying to capture presence and sentiment instead of categorizing into a 226 topic classification system.

Our work is systemically different from previous computer context analysis work that political scientists are familiar with. Our method uses a case-based or "learning by example" approach to replicate the ability of humans to classify individual documents. The software trains on human labeled examples before proceeding to label other documents at high levels of accuracy. Other researchers have

developed useful information extraction methods for content analyzing political documents. KEDS/TABARI (Schrodtt, Davis, and Weddle, 1994), King and Lowe (2003), and Wordscores (Laver and Garry, 2003), select 'cases' for learning-by-example in distinctly different ways. With KEDs/TABARI and similar methods, humans create the rules that the software uses to code events. Wordscores uses reference document cases to rank subsequent documents. In our work, humans assign a small number of cases to mutually exclusive topic categories. Using these initial cases as a training set, the software then computationally generates rules for the correct labeling of unlabeled cases. Once the rules have been generated, users feed unlabeled cases to the system for categorization. To test performance, we assess whether the system's predictions for the unlabeled cases accord with those assigned by human coders. To ensure that our methods aren't unacceptably biased due to agreement-by-chance or high variance within specific categories, our performance metrics discount performance assessment for these factors. For these reasons and others, our methodology is considerably different from the above mentioned approaches as well as unsupervised approaches to detecting patterns in data (Quinn et al., 2006). In computational linguistics, unsupervised approaches are frequently used as a first step toward developing topic classification systems in the absence of strong priors about the organization of such a system (Hand, Mannila, & Smyth, 2001).

Like our other work, we are still interested in lowering the costs of coding contemporary data. And we still need to deal with problems of topic drift over time; because today's political issues are not a representative sample of the issues discussed in 2001. The language of web sites changes with the topics over time as hot issues on the agenda shift. An efficient automated system must be able to recognize and adapt to such changes in framing dynamically. This work demonstrates our first successful systemic attempts in this regard.

The machine learning algorithms used here are more similar to the system presented in Purpura & Hillard (2006) than it is to the one used in Hillard, Purpura, and Wilkerson (2007). But we have modified the features used to examine the data to include presence and discount for temporal decay<sup>9</sup>.

---

<sup>9</sup> This method will be described in more detail in a future version of this paper. The temporal inductive transfer experiments were just recently completed. The method is based upon the notion that the probability of the word selection which discriminate a topic can be estimated without knowledge of the classes of the topic, and that these probabilities change from month to month. The simplest way to think about it is that the software estimates the probability that 'choice' has lost discriminative power in a topic category when it diminishes in importance in the time series corpus. When it reappears, we associate the word with its new topic instead of the previous usage.

Once again, we build on existing automatic text classification research in the Computer Science and Computational Linguistics literatures. A relatively comprehensive analysis (Yang and Liu, 1999) finds that support vector machines are usually the best performing model. The most typical feature representation first applies Porter stemming to reduce word variants to a common form (Porter, 1980) before computing term frequency in a sample divided by the inverse document frequency (to capture how often a word occurs across all documents in the corpus) (Papineni, 2001). A list of common words (stop words) may also be omitted from each text sample.

## **Machine Learning Methods**

The main purpose of automated text classification is to replicate and assist the performance of human labelers. In this case, the classification task consists of 4 categories, positive, neutral, negative, and not found. We use SVMlight for SVM classification (Joachims, 1998) as the toolkit to support our work.

The SVM system builds on binary one vs. one classifiers between each pair of categories, and selects a final category by choosing the category that is selected most often by the one vs. one classifiers. Other approaches are also common (such as a committee of one vs. all classifiers), but we found our approach to be more time efficient, with equal or greater performance. We use a linear kernel, Porter stemming, and a feature value that is slightly more detailed than the typical inverse document frequency feature. In addition, we prune words in each web site that occur less often than the corpus average. Further details and previous results of the system are described in (Purpura and Hillard, 2006). For this project, we have also added additional features which specify temporal decay through an effective expansion of vocabulary to discount the appearance of words which fail to always appear frequently within a category. In addition to these changes, we have added binary presence indicators. The addition of temporal inductive systems now requires each web site to have a time stamp.

## **Evaluation Methods and Metrics**

We check human coders and the computer 'coder' in the same way. Evaluation of success is based on agreement among the humans and the machine. Unlike our previous work, this project does not have a high quality gold standard of meticulously analyzed code samples. Our coding system was developed ad-hoc to answer questions primarily for this research. But we still assess performance against human labelers using inter-coder statistics and by checking the face validity of analysis results. In addition, since our methods label every web site, we can go directly to the source. Later in the paper, we will offer examples of this.

When we measure inter-coder reliability, instead of simply counting the number of records which have agreement between the machine generated predictions and the human team’s estimate, our analysis methods discount performance for confusion and for the probability of agreement by chance. In our case, the probability of agreement by chance is quite high. But our methods discount inter-coder agreement performance for this factor. Our methods also discounts for small numbers of records, another problem which isn’t a significant factor in our analysis.

We use inter-coder agreement statistics common in the topic classification literature to easily enable comparison with other systems. For this reason, we report Cohen’s Kappa (Cohen, 1968) and AC1 (Gwet, 2002) to assess inter-coder agreement with the human team.

Cohen’s Kappa statistic is a standard metric used to assess inter-coder reliability between two sets of results. Usually, the technique is used to assess results between two human coders, but the computational linguistic field uses the metric as a standard mechanism to assess agreement between a human and machine coder. Cohen’s Kappa statistic is defined as:

$$\kappa = \frac{p(A) - p(E)}{1 - p(E)} \quad (1)$$

In the equation,  $p(A)$  is the probability of the observed agreement between the two assessments:

$$p(A) = \frac{1}{N} \sum_{n=1}^N I(\text{Human}_n == \text{Computer}_n) \quad (2)$$

Where  $N$  is the number of examples, and  $I()$  is an indicator function that is equal to one when the two annotations (human and computer) agree on a particular example.  $P(E)$  is the probability of the agreement expected by chance:

$$p(E) = \frac{1}{N^2} \sum_{c=1}^C (\text{HumanTotal}_c \times \text{ComputerTotal}_c) \quad (3)$$

Where  $N$  is again the total number of examples and the argument of the sum is a multiplication of the marginal totals for each category. For example, for category 3, health, the argument would be the total number of bills a human coder marked as category 3, times the total number of bills the computer system marked as category 3. This multiplication is computed for each category, summed, and then normalized by  $N^2$ .

Due to bias in certain case conditions (Gwet, 2002), computational linguists also use another standard metric named the AC1 statistic to assess inter-coder reliability. The AC1 statistic corrects for the bias of Cohen’s Kappa by calculating the agreement by chance in a different manner. It has similar form:

$$AC1 = \frac{p(A) - p(E)}{1 - p(E)} \quad (6)$$

But the  $p(E)$  component is calculated differently:

$$p(E) = \frac{1}{C-1} \sum_{c=1}^C (\pi_c (1 - \pi_c)) \quad (7)$$

Where  $C$  is the number of categories, and  $\pi_c$  is the approximate chance that a bill is classified as category  $c$ .

$$\pi_c = \frac{(\text{HumanTotal}_c + \text{ComputerTotal}_c) / 2}{N} \quad (8)$$

For coding problems of sentiment analysis, a Cohen's Kappa or AC1 statistic of 0.40 or higher is considered to be very good agreement between coders.

## Data and Experiments

Testing our propositions is a two stage process. First, we gather the data and then we examine it to determine whether our propositions are observable or not. At this time, we have not finished analyzing all of the data. For data analysis simplicity, we restrict our analysis to quarters of the year from 2001 to 2006. Our data set consists of records for each Congressional Member for each quarter from Q3 2001 to Q4 2006. In each record, is MemberID, In Congress (binary), Party, Presidential Approval at the end of the Quarter, 2004 Bush Vote Share, 2000 Bush Vote Share, Vulnerable (binary), Web Site Presence (binary), Positive Mention of President Bush (binary), Negative Mention of President Bush (binary), Neutral Mention of President Bush (binary), and a few other district demographic details which are unnecessary for this analysis. This results in 9,570 records.

Member web site snapshots are not available for every day of the week. We have captured many of the web sites from Archive.org and we also captured many snapshots directly from the Congressional sites themselves. Figure 1 shows a listing of some of the dates that we captured snapshots for Senator Akaka from Hawaii. To reduce complexity for this paper's analysis, we converted individual snapshot records to quarterly records, by ORing binary variables across all of the observations in a quarter.

**Figure 1: List of web site snapshots available from Archive.org for Senator Daniel Akaka (D-HI)**

	1999	2000	2001	2002	2003	2004	2005	2006
s	2 pages	6 pages	74 pages	15 pages	20 pages	27 pages	64 pages	16 pages
	<a href="#">Oct 19, 1999</a> * <a href="#">Dec 18, 1999</a>	<a href="#">Mar 05, 2000</a> <a href="#">Jun 22, 2000</a> * <a href="#">Aug 15, 2000</a> * <a href="#">Aug 21, 2000</a> <a href="#">Oct 18, 2000</a> <a href="#">Dec 04, 2000</a> *	<a href="#">Feb 02, 2001</a> <a href="#">Feb 05, 2001</a> <a href="#">Mar 01, 2001</a> <a href="#">Mar 02, 2001</a> <a href="#">Mar 03, 2001</a> <a href="#">Apr 01, 2001</a> <a href="#">Apr 05, 2001</a> <a href="#">Apr 08, 2001</a> <a href="#">Apr 14, 2001</a> <a href="#">May 04, 2001</a> <a href="#">Jul 20, 2001</a> <a href="#">Sep 13, 2001</a> <a href="#">Sep 14, 2001</a> <a href="#">Sep 15, 2001</a> <a href="#">Sep 16, 2001</a> <a href="#">Sep 17, 2001</a> <a href="#">Sep 18, 2001</a> <a href="#">Sep 20, 2001</a> <a href="#">Sep 21, 2001</a> <a href="#">Sep 23, 2001</a> <a href="#">Sep 24, 2001</a> <a href="#">Sep 25, 2001</a> <a href="#">Sep 27, 2001</a> * <a href="#">Sep 28, 2001</a> <a href="#">Sep 29, 2001</a> <a href="#">Sep 30, 2001</a> <a href="#">Oct 01, 2001</a> <a href="#">Oct 02, 2001</a>	<a href="#">Jan 20, 2002</a> <a href="#">Feb 28, 2002</a> <a href="#">Mar 24, 2002</a> <a href="#">Mar 29, 2002</a> <a href="#">Apr 06, 2002</a> <a href="#">May 28, 2002</a> <a href="#">Jul 23, 2002</a> <a href="#">Aug 02, 2002</a> <a href="#">Aug 06, 2002</a> <a href="#">Sep 24, 2002</a> <a href="#">Oct 05, 2002</a> <a href="#">Oct 10, 2002</a> <a href="#">Oct 18, 2002</a> <a href="#">Nov 21, 2002</a> * <a href="#">Dec 30, 2002</a>	<a href="#">Feb 04, 2003</a> <a href="#">Feb 18, 2003</a> <a href="#">Mar 22, 2003</a> * <a href="#">Apr 04, 2003</a> <a href="#">Apr 24, 2003</a> <a href="#">May 13, 2003</a> <a href="#">May 24, 2003</a> <a href="#">Jun 02, 2003</a> <a href="#">Jun 04, 2003</a> <a href="#">Jun 17, 2003</a> <a href="#">Jul 24, 2003</a> <a href="#">Aug 05, 2003</a> <a href="#">Sep 23, 2003</a> <a href="#">Oct 15, 2003</a> * <a href="#">Oct 18, 2003</a> <a href="#">Oct 25, 2003</a> <a href="#">Nov 28, 2003</a> <a href="#">Dec 04, 2003</a> <a href="#">Dec 06, 2003</a> <a href="#">Dec 20, 2003</a>	<a href="#">Feb 03, 2004</a> <a href="#">Mar 26, 2004</a> <a href="#">Apr 02, 2004</a> <a href="#">Jun 05, 2004</a> <a href="#">Jun 09, 2004</a> <a href="#">Jun 12, 2004</a> <a href="#">Jun 15, 2004</a> <a href="#">Jun 16, 2004</a> <a href="#">Jul 18, 2004</a> * <a href="#">Jul 24, 2004</a> <a href="#">Jul 28, 2004</a> <a href="#">Aug 05, 2004</a> <a href="#">Aug 21, 2004</a> <a href="#">Aug 30, 2004</a> <a href="#">Sep 12, 2004</a> <a href="#">Sep 20, 2004</a> <a href="#">Sep 30, 2004</a> <a href="#">Oct 15, 2004</a> <a href="#">Oct 24, 2004</a> <a href="#">Nov 11, 2004</a> <a href="#">Nov 15, 2004</a> <a href="#">Nov 17, 2004</a> <a href="#">Nov 20, 2004</a> <a href="#">Nov 28, 2004</a> <a href="#">Nov 30, 2004</a> <a href="#">Dec 05, 2004</a> <a href="#">Dec 16, 2004</a>	<a href="#">Jan 26, 2005</a> <a href="#">Jan 31, 2005</a> <a href="#">Feb 04, 2005</a> <a href="#">Feb 05, 2005</a> <a href="#">Feb 06, 2005</a> <a href="#">Feb 08, 2005</a> <a href="#">Feb 15, 2005</a> <a href="#">Feb 16, 2005</a> <a href="#">Mar 03, 2005</a> * <a href="#">Mar 05, 2005</a> <a href="#">Mar 08, 2005</a> <a href="#">Mar 09, 2005</a> <a href="#">Mar 10, 2005</a> <a href="#">Mar 14, 2005</a> <a href="#">Mar 20, 2005</a> <a href="#">Mar 23, 2005</a> <a href="#">Mar 28, 2005</a> <a href="#">Mar 31, 2005</a> <a href="#">Apr 01, 2005</a> <a href="#">Apr 05, 2005</a> <a href="#">Apr 08, 2005</a> <a href="#">Apr 23, 2005</a> <a href="#">May 14, 2005</a> * <a href="#">May 24, 2005</a> * <a href="#">Jun 01, 2005</a> <a href="#">Jun 02, 2005</a> <a href="#">Jun 17, 2005</a> <a href="#">Jun 18, 2005</a>	<a href="#">Jan 03, 2006</a> <a href="#">Jan 18, 2006</a> <a href="#">Jan 25, 2006</a> * <a href="#">Jan 27, 2006</a> * <a href="#">Jan 28, 2006</a> <a href="#">Jan 30, 2006</a> <a href="#">Feb 03, 2006</a> <a href="#">Feb 08, 2006</a> <a href="#">Feb 14, 2006</a> <a href="#">Feb 21, 2006</a> <a href="#">Mar 16, 2006</a> <a href="#">Mar 17, 2006</a> <a href="#">Mar 19, 2006</a> <a href="#">Mar 25, 2006</a> <a href="#">Mar 28, 2006</a> <a href="#">Mar 29, 2006</a>

We label Republicans as “vulnerable” if the President had a lower share of the vote in their district in 2004 or if the Cook Report listed them as vulnerable prior to the election. All other incumbent Republicans are invulnerable. We use a logically consistent version of the process for Democrats.

Our data set only examines active members of Congress in 2006. If a Member was elected in 2004, we exclude all of their irrelevant records.<sup>10</sup> We calculate inter-coder reliability statistics for the data that we have labeled and then prepare a series of analyses to examine our propositions.

<sup>10</sup> There are many sophisticated modeling approaches we can take to examine the panel data that we have. For the short term, we have chosen the simplest methods. We hope you will see shortly that the patterns are obvious and glaring, given the type of conclusions we are attempting to draw.

## Inter-coder Reliability

For nearly 25,000 records, we calculated inter-coder reliability across 2 humans and the computer. One of the humans coded very few records, while the other did many. Some of their work overlapped but not all of it. The results in Table 1 indicate that the humans had a high level of inter-coder agreement between each other, and very good agreement with the computer on presence. In addition, the sentiment scores are considered very good for the complexity of the task. The primary problem in sentiment comparison was differentiating between Positive and Neutral. Interestingly, one of the human coders had very high agreement with the machine for measuring sentiment across a low number of observations. Note that an inter-coder reliability score of 0.54 is not slightly better than flipping a coin. It is significantly better than flipping a coin because the measure discounts for agreement by chance.

<b>Table 1: Inter-coder Reliability for Human and Computer Coders (AC1)</b>				
	<b>Human 1 vs. Human 2</b>	<b>Human 1 vs. Computer</b>	<b>Human 2 vs. Computer</b>	<b>Humans vs. Computer</b>
<b>Presence 2 categories</b>	0.96	0.85	0.86	0.85
<b>Sentiment 3 categories</b>	0.89	0.51	0.72	0.54
Results are based on using 24,950 records. Human 2 coded only 1100 records.				

In addition to the inter-coder reliability tests, basic assessment of the validity of the snapshots was conducted by looking at some of the examples. The system gives the ability to display screenshots, as demonstrated below:

# Snapshot 1: Burns, 2-12-05



Home Biography Student Corner Services Issues Committees News Center Events Contact Info Links Let Me Help You

## Featured Story



### Burns Disappointed in Education Cuts

Sends Letter to Budget Chair Outlining Concerns

WASHINGTON-U.S. Sen. Conrad Burns (R-Mont.) wrote a letter to Budget Committee Chairman Judd Gregg (R-NH) this week in response to the education cuts included in the President's proposed budget. "I am extremely disappointed that the budget did not include funding for critical vocational education programs," said Burns. "We also saw cuts to Federal TRIO and other high school-related programs that work to decrease Montana's high school drop out rates, despite their effectiveness. These programs make a difference in the lives of Montana's young people and I will continue to work hard to ensure that these successful programs continue."

Welcome to my website. Please feel free to contact me with any thoughts and concerns that you may have. [\[Video Greeting\]](#)

 [EMAIL ME](#)

**eNewsletter:**  
Sign up here!  


**Quick Search:**  
  


**Contact Info:**  
Washington Office  
187 Dirksen Senate  
Office Building  
Washington D.C.  
20510  
(202) 224-2644  
(202) 224-8594 fax  
(202) 224-8616 TDD  
1-800-344-1513

Burns has supported various vocational initiatives in the past, including the Carl D. Perkins Career and Technical Education Improvement Act, which provides more coordination and flexibility to ensure st... [\[ more \]](#)



[Request A Flag](#)



[D.C. Tours](#)



[Service Academy](#)



[Field Offices](#)



[Let me Help You](#)



[Pages/Interns](#)

[My Privacy Statement](#)

#### Latest News

- 02-10 Burns Disappointed in Education Cuts
- 02-10 U.S. will not accept meat from older Canadian cows
- 02-10 Burns raps Corps over hatchery funding
- 02-09 Burns Announces e-Eleven Technology Agenda
- 02-09 Burns Announces Fire Safety and Prevention Funding for Melrose, Polaris and Sidney

[\[View All News\]](#)

#### Latest Events

- 01-27 Community Outreach Hours In Lewistown
- 01-27 Community Outreach Hours in Deer Lodge
- 01-26 Community Outreach Hours in Plains and Thompson Falls
- 12-15 Burns Rep To Hold Hours In Choteau

[\[View All Events\]](#)

**NEW MAIL POLICY**

**Support the Troops**



## Snapshot 2: Capito, 3-25-05

U.S. Congresswoman  
**SHELLEY MOORE CAPITO**  
Representing West Virginia's 2nd Congressional District

[Write Shelley](#)  
[Email Sign Up](#)  
[Visiting Washington, DC?](#)

[About Shelley](#) [District Information](#) [Constituent Information](#) [Legislative Issues](#) [Press Office](#)



### Capito Secures \$44 Million For U.S. Route 35

**Record Funding Puts Construction Back Into Gear**

**WASHINGTON, DC - (3/10)** Over \$44 million of long awaited funding from the federal Highway Trust Fund to build the new U.S. Route 35 in Mason and Putnam Counties was secured by Congresswoman Shelley Moore Capito (R-WV) as part of a massive highway funding bill currently under debate by the U.S. House of Representatives she noted today. ([Read More](#))



- [U.S. Route 35 Photo Gallery](#)

### Proposed Legislation Would Allow Coal Trucks to Avoid Chesapeake



A coal truck rolls through downtown Chesapeake, WV on Route 61.

### Capito Voices Concerns with Presidents Budget Proposal

**WASHINGTON, DC - (3/3)** During testimony before the House Budget Committee, Congresswoman Shelley Moore Capito (R-WV) expressed her concern with President Bush's Budget proposal telling the committee, "...funding cuts will do substantial harm to economic development activities in rural areas." ([Read More](#))

### Recent News

Capito Supports Terri Schiavos Right to Have Her Day in Federal Court

Capito Urges Senate Action on Measures to Lower Energy Costs

Capito Leads Debate against Indecency on TV

Capito Urges More Women to Enter Public Service

Capito Dear Colleague Stresses Importance of Civility in the House of Representatives

Death Gratuity Proposal Good First Step, DoD Should Expand Efforts to Include all Soldiers


### Weekly Column

Is Your Gasoline "Made in the USA?"

**Snapshot 3: Chafee, 3-3-05. This criticism of the nomination of Gonzales to the Attorney General is difficult to read, but negative.**

<p><b>Chafee Statement on Alberto Gonzales Nomination to be Attorney General:</b> "The Administration adopted an aggressive legal strategy in the aftermath of the attacks of September 11th, 2001 as it sought to prevent future attacks and ascertain the scope of the threat facing our nation. It is now clear that the Administration badly overreached in certain areas, and I am saddened and angered by allegations of abuse of foreign detainees which have done significant damage to our image throughout the world. (2/3/05) <a href="#">Read More »</a>"</p> <p><b>Statement of Senator Chafee on the Success of Elections in Iraq:</b> "The historic elections held in Iraq demonstrate the yearning of human beings to participate in their government even under duress and danger. I applaud our troops as well as the election officials for improving the security situation across Iraq, allowing for these elections to take place." (1/31/05)</p> <p><a href="#">Latest Photos and Press Releases</a></p>	<p>Senator Chafee poses with Senator Elizabeth Dole and Warwick Mayor Scott Avedisian in his office on the morning of the Inauguration of President George W. Bush. Chafee hosted a breakfast for Rhode Islanders attending the ceremonies.</p> <p><a href="#">Click Here to View More Inauguration Pictures »</a></p>
---	--

**Snapshot 4: Fossella , 5-30-05. Criticism of Social Security.**

	<p><a href="#">Latest News</a></p> <h2>Headlines</h2>
<p>5/27/2005-<a href="#">Rep. Fossella Announces Opposition to Bush Social Security Proposal</a></p> <p><i>[Washington, DC] Congressman Vito Fossella (R-NY13) today announced his opposition to a key element of President Bush's Social Security plan, declaring that progressive indexing would hurt many residents in Brooklyn.</i></p>	

**Testing the System**

Our first test of the system examined the behavior of Congressional Republicans after September 11<sup>th</sup>, 2001 and just before the election in 2006. This presence test involved a simple a check of whether the data seemed reasonable. The findings for Republicans are listed in Table 2. The results are stark and were double-checked by human coders to verify accuracy. Near the 2006 election, only four vulnerable Republicans would even mention the President on their web site and they all criticized him. The valence of the depiction of the President appeared strongly associated with the broad popularity of the President among constituents.

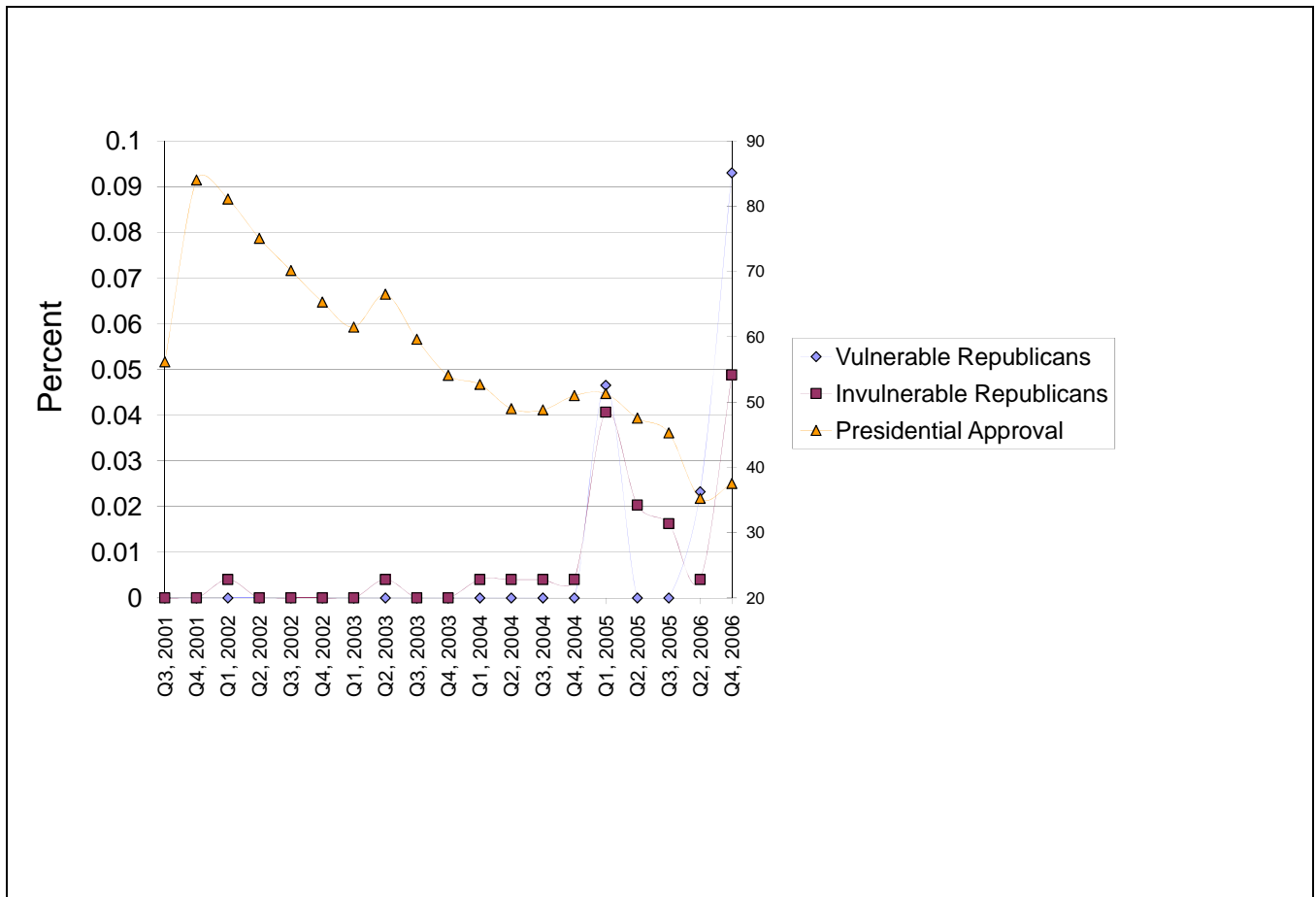
<b>Table 2: Presence of Mentions of the President Panel State Analysis of Vulnerable and Safe Republicans</b>			
	<b>Vulnerable</b>	<b>Safe</b>	<b>Presidential Approval</b>
<b>Q4 2001</b>	66.67%	35.29%	90%
<b>Q4 2006</b>	1.75%	14.04%	37%
Results are based on 9,570 records			

### **Considering the Republican Behavior in Context**

The Republican behavior is even more interesting when the trend is considered. Figure 2 shows that Republicans maintained strict message discipline as the President's popularity declined except when the President's budget and social security plan were announced in the State of the Union in 2005. At that point, safe Republicans broke with the President over domestic policy. The party eventually went back into line and then broke again when the 2006 election cycle began. After Q2, 2006 the negative mentions of the President's policy dominated the web site mentions by both Republicans and Democrats.

More interesting from a methodological perspective, there is no significant difference in the trend lines generated by computer coding or human coding for this graph. Despite fairly low levels of inter-coder reliability, since most of the errors are related to differentiating positive and neutral mentions, the negative mention graph doesn't significantly change.

**Figure 2: Negative Mentions of President Bush's Policies Compared to Presidential Approval (Gallup). The orange line with diamond indicators is Presidential approval. Republicans maintained strict message discipline, rarely criticizing the President until Q3 2006. A brief blip in Q1 2005 is the study of a case study later in the paper. Vulnerable Republicans were more likely to criticize the President as the President's approval declined below 40% and the 2006 elections approached.**



## Discussion

This work is in its infancy – we just completed analysis of a small part of the data. We've examined it with logit and the differences in probability are significant, but we want to conduct additional testing before we make any broader claims than the following.

In many ways the results that we see are expected. Our propositions hold. Proposition 1: The valence of the depiction of the President will be strongly associated with the broad popularity of the President among constituents seems true. Although the data appears to be heavily influenced by events on dates, popularity of the President seems to play some role. The question is whether events play a larger role and we need to find a mechanism to test whether it is really events driving the picture.

We further asserted that strategic calculations are always being made and that it is most important when electoral considerations are pre-eminent, specifically for vulnerable incumbents near an election. The evidence seems to suggest this, but, again, events may be driving the story and we need to better test for that.

In terms of Proposition 2: The strategic calculus as to how to depict relationship to the President is heightened with the proximity to the election and the vulnerability of the incumbent. This also appears to be true, but again, events may be more of a driver and we wait for better models to improve our confidence in the solution.

The one thing that we can take away from this experiment is that the software tool is incredibly valuable. In addition to classifying documents for sentiment, it enables fast, highly accurate search and retrieval for documents of interest. You can drive directly from statistical quantities of interest or outliers into the cases themselves.

## References

- Forman, G. June 20, 2006. "Tackling Concept Drift by Temporal Inductive Transfer". Research Note: HP Labs.
- Gwet, K. Kappa Statistic is not Satisfactory for Assessing the Extent of Agreement Between Raters. in *Statistical Methods For Inter-Rater Reliability Assessment, No. 1*, April, 2002.
- Jacobson, Gary. 2004. *The Politics of Congressional Elections*. Sixth Edition. NY: Pearson.
- Adler, E. Scott and John D. Wilkerson. (forthcoming) "Intended Consequences? Committee Reform and Jurisdictional Change in the House of Representatives." *Legislative Studies Quarterly*.
- Baum, M. *Soft News Goes to War: Public Opinion and American Foreign Policy in the New Media Age*. 2003. Princeton N.J.: Princeton University Press.
- Baumgartner, Frank, Bryan Jones, and John Wilkerson. 2002. "Studying Policy Dynamics." In Baumgartner and Jones, *Policy Dynamics*, Chapter 2.
- Eric Brill and Jun Wu. 1998. "Classifier combination for improved lexical disambiguation". In *Proc. ACL*, pages 191–195.
- Breiman, L. 2001. "Statistical Modeling: The Two Cultures". *Statistical Science*. 16(3):199-231.
- Cristianini, N., Shawe-Taylor, J., and Lodhi, H. Latent semantic kernels. in Brodley, C. and Danyluk, A. Proceedings of ICML-01, *18th International Conference on Machine Learning*. (San Francisco, US, 2001), Morgan Kaufmann Publishers, pages 66–73.
- Cohen, J. 1968. "Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit." *Psychological Bulletin*, 70(4):213–220.
- Cohn, D. and L. Atlas, and R. Ladner. 1994. "Improving generalization with active learning". *Machine Learning*, 315:201–221.
- Curran, J. 2002. "Ensemble methods for automatic thesaurus extraction". *Proc. Empirical Methods in Natural Language Processing*, pages 222–229.
- Dagan, I. and S. Engelson. 1995. "Committee-based sampling for training probabilistic classifiers". *Proc. Int. Conf. on Machine Learning*, pages 150–157.
- Dekel, O., J. Keshet, and Y. Singer. 2004. "Large margin hierarchical classification". *ACM International Conference Proceeding Series*.
- Dietterich, T. 2000. "Ensemble methods in machine learning". *Lecture Notes in Computer Science*, 1857:1–15.

- Fowler, James H. 2006. "Legislative cosponsorship networks in the US House and Senate." *Social Networks* 28: 454-465.
- Freund, Y., H.S. Seung, E. Shamir, and N. Tishby. 1997. "Selective Sampling Using the Query by Committee Algorithm". *Machine Learning*, 28(2):133–168.
- Gwet, K. "Kappa Statistic is not Satisfactory for Assessing the Extent of Agreement Between Raters." in *Statistical Methods For Inter-Rater Reliability Assessment, No. 1*, April, 2002.
- Hand, D., Mannila, H., Smyth, P. 2001. *Principles of Data Mining (Adaptive Computation and Machine Learning)*. Cambridge: The MIT Press.
- Hand, D. 2006. "Classifier Technology and the Illusion of Progress". *Statistical Science*. 21(1):1–14.
- Joachims, T. 1998. "Text categorization with support vector machines: Learning with many relevant features". In Proc. *European Conference on Machine Learning*.
- King, D. 1997. *Turf Wars*. Chicago: University of Chicago Press.
- King, G. and Lowe, W. 2003. "An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design." *International Organization* 57(3, July):617–642.
- King, G. R. O. Keohane, S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Koller, D. and M. Sahami. 1997. "Hierarchically classifying documents using very few words". Proc. Int. Conf. on Machine Learning, pages 170–178.
- Lewis, D. and J. Catlett. 1994. "Heterogeneous uncertainty sampling for supervised learning". In Proc. Int. Conf. on Machine Learning.
- Liere, R. and P. Tadepalli. 1997. "Active learning with committees for text categorization". Proceedings of the Fourteenth National Conference on Artificial Intelligence, pages 591–596.
- Mann, G., Mimno, D. and McCallum, A. 2006. "Bibliometric impact measures leveraging topic analysis". JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries. Pp 65–74. Chapel Hill, NC, USA.
- McCallum, A., R. Rosenfeld, T. Mitchell, and A. Ng. 1998. "Improving text classification by shrinkage in a hierarchy of classes". Proc. Int. Conf. on Machine Learning, 367.
- McCallum, A. 1996. "Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering". <http://www.cs.cmu.edu/mccallum/bow>.
- Muslea, I. 2006. "Active learning with multiple views". *Artificial Intelligence Research*, 27:203–233.

- Papineni, K. 2001. "Why inverse document frequency?" In Proc. North American Chapter of the Association for Computational Linguistics.
- Pollack, H. and R. Zeckhauser. 1996. "Budgets as dynamic gatekeepers". *Management Science*, 42:642–658.
- Porter, M.F. 1980. "An algorithm for suffix stripping". *Program*, 16(3):130–137.
- Purpura, S and Hillard, D. 2006. "Automated classification of congressional legislation". In Proc. Digital Government Research, pages 219–225.
- Przeworski, A and Teune, H. 1970. *The Logic of Social Inquiry*. New York: Krieger Pub Co (March 1982).
- Quinn, K, Monroe, B, Colaresi, M., Crespin, M. and Radev, D. 2006. "An Automated Method of Topic-Coding Legislative Speech Over Time with Application to the 105th-108th US Senate." Annual Meetings of the Society for Political Methodology.
- Schrodt P., Davis, S., and Weddle, J. 1994. "Political Science: KEDS—A Program for the Machine Coding of Event Data". *Social Science Computer Review*, Vol. 12, No. 4, 561-587 (1994).
- Schapire, R.E. and Y. Singer. 2000. "Boostexter: A boosting based system for text categorization". *Machine Learning*, 39(2/3):135–168.
- [Schrodt, P and D. Gerner. 1994. "Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982-92" \*American Journal of Political Science\*. 38 \(3\): 825-44/](#)
- Sebastiani, F. "Machine learning in automated text categorization". *ACM Computing Surveys*, 34(1).
- Soroka, S., C. Wlezien, and I. McLean. 2006. "Public Expenditure in the UK: How Measures Matter". *Journal of the Royal Statistical Society*, pages 255–271.
- Sugiyama, M. and H. Ogawa. 1999. "Functional analytic approach to model selection—Subspace information criterion". Proc. Workshop on Information-Based Induction Sciences, pages 93–98.
- Yang, Y. and X. Liu. 1999. "A re-examination of text categorization methods". In Proc. of SIGIR-99.
- Yang, H., Callan, J., and Shulman, S. 2006. "Next steps in near-duplicate detection for eRulemaking." *Proceedings of the Sixth National Conference on Digital Government Research*. San Diego, CA.
- [Identifying Expressions of Opinion in Context](#). Eric Breck, Yejin Choi, and Claire Cardie. To appear at the *Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.
- [Joint Extraction of Entities and Relations for Opinion Recognition](#). Yejin Choi, Eric Breck, and Claire Cardie. *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2006.

[Partially Supervised Coreference Resolution for Opinion Summarization through Structured Rule Learning](#). Veselin Stoyanov and Claire Cardie. *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2006.

[Toward Opinion Summarization: Linking the Sources](#). Veselin Stoyanov and Claire Cardie. *COLING\ACL2006 Workshop on Sentiment and Subjectivity in Text*, 2006.

[Annotating Expressions of Opinions and Emotions in Language](#). Janyce Wiebe, Theresa Wilson, Claire Cardie. *Language Resources and Evaluation (formerly Computers and the Humanities)*, 39:2-3, 2005.

[Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns](#). Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. *Proceedings of HLT-EMNLP 2005*, 2005.