

PSST: A Web-Based System for Tracking Political Statements

Samantha Kleinberg, Bud Mishra
 Courant Institute of Mathematical Sciences, New York University
 New York, NY, USA
 samantha@cs.nyu.edu, mishra@nyu.edu

ABSTRACT

Determining candidates' views on important issues is critical in deciding whom to support and vote for; but finding their statements and votes on an issue can be laborious. In this paper we present PSST, (Political Statement and Support Tracker), a search engine to facilitate analysis of political statements and votes over time. We show that prior tools for text analysis can be combined with minimal manual processing to provide a first step in the full automation of this process.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval; J.4 [Computer Applications]: Social and Behavioral Sciences

General Terms

Algorithms

1. INTRODUCTION

During the 2004 US Presidential election, the notion of “flip-flopping” was made salient, but despite analyses[2, 3, 4] of speeches given by both candidates in the final months before the election, the “flip-flopper” label was not applied equally. In this work, we aim to provide a non-partisan, unbiased, method of viewing how public statements on issues change over time as well as how statements correlate with the politicians' votes on these issues. This is useful for voters who value consistency and want to be aware of candidates' histories, as well as for holding politicians accountable for their statement histories both before and after elections. Ideally, if this could be done in an automated way, for any politician or candidate, as the news is happening, any person could check the facts for themselves immediately - rather than having just a feeling that rhetoric has changed. On the other side of things, if politicians are aware that anyone can check on their histories in a simple way, we may avoid this shift in arguments and rationale. We present our first work toward this end, PSST (Political Statement and Support Tracker), a web-based system currently allowing analysis many of the candidates in the 2008 United States Presidential election, and President Bush, on a set of issues and key votes.

Copyright is held by the author/owner(s).
 WWW 2008, April 21–25, 2008, Beijing, China.
 ACM 978-1-60558-085-2/08/04.

2. METHODS

The main steps of PSST are: (1) Retrieving webpages with speeches, press releases, and votes (2) Extracting relevant quotes from the texts and relating these to political issues (3) Displaying the statements and voting histories.

2.1 Data collection

In order to analyze candidates' consistency on issues we gathered speech transcripts from their official websites, and later examined these to determine their focus and find statements on particular issues. Voting records were also gathered and similarly analyzed. By using the politicians' own words, we obtain a truer representation of their publicly stated positions, not filtered through the lens of the media.

As valid RSS feeds of speeches do not exist for all candidates, and news stories proved to be too ambiguous, we restricted the system to a pre-defined subset of all political figures. The candidates and politicians currently supported in the PSST system are: George Bush, Hillary Clinton, John Edwards, John McCain, Barack Obama, Mitt Romney, and Fred Thompson. These were the top candidates, according to polls at the time, for which speech transcripts were available. A full list of data sources is given on the project web site. Similarly, for politicians with voting histories, key votes were retrieved from the Project Vote Smart website.¹

2.2 Getting key phrases

In the second phase of PSST, key phrases are extracted from the texts and then linked with issues. Here Extractor[1][5] was used for key phrase extraction, as well as for identifying the supporting text of those key phrases.

Given a link to a speech, Extractor returns a list of the key phrases in the speech and, for each phrase, a list of the sentences in the text that support that phrase. These phrases represent the topic of the text. For example, a speech about health care policy may return phrases such as “insurance companies” and “health.” Other techniques for identifying key phrases, such as TF-IDF, tend to focus on unusual or frequently used words that do not always accurately describe the speech's focus.

PSST compares the key phrases found by Extractor with a list of words and phrases that we have manually identified as relating to specified political issues (e.g. campaign finance, the environment, etc.). For each occurrence of such a phrase in the speech, PSST searches the surrounding sentences for any additional significant phrases that might indicate other relevant issues. (See website for full list of issues supported).

¹www.vote-smart.org

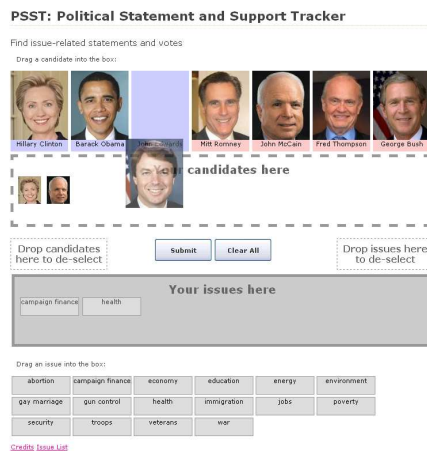


Figure 1: Query interface in use. The user is adding a third candidate to the selected query.

Then, using the same rules as for the statements we automatically classified each of the bills.

In order to determine these main issue categories as well as what words fall within them, we studied websites such as that of Project Vote Smart to get an idea of the general categories the statements fell into and then reviewed the phrases and sentences extracted for one Democratic and one Republican candidate (Barack Obama and John McCain) to see which phrases correspond to which issues and account for different wordings between parties. Manually reviewing the phrases and sentences extracted for each speech aided in finding issue specific language. For each speech we looked at the phrases extracted, verified that they corresponded to at least one of the points of the speech and then determined based on common and issue specific knowledge what category the phrase belonged in, and what other similar phrases may be used to make such classifications. For example, phrases such as “contraception” and “family planning” are common in texts dealing with abortion but rare in other contexts. We continued this process using vote titles, in order to identify more phrases that had originally been missed or that might only be used in the context of a bill rather than a speech. One example of this is CHIP (Child Health Insurance Plan), as there were a number of votes on the plan, though it was rarely mentioned by name outside of bill titles.

2.3 Queries

At query time, results are retrieved from the database. Using a drag and drop interface, a user can select any number of candidates and issues. Then, for each issue and each candidate selected, the related votes and statements are retrieved and arranged in descending chronological order with votes/statements displayed together. Figure 1 shows the user interface.

3. DISCUSSION

The system is successful in aiding the discovery of inconsistencies between statements and votes, highlighting the reasoning for a vote that may seem inconsistent with a politician’s stated beliefs, and facilitating the identification of

changing positions in cases where votes are not available. Rather than simply doing a search on Google for a politician or issue, a user can immediately find all statements for multiple candidates on the topics that interest them. We were able to identify inconsistencies such as changes in President Bush’s arguments for war with Iraq, Senator Clinton’s statements on bringing troops home versus her votes against redeployment, and Senator McCain’s rationale for a vote against a health care bill after stating the importance of health care reform.

However, it does not succeed in all areas. In some cases, particularly with categorizing votes based solely on their titles, there were misclassifications. For example “Healthy Forests Restoration Act of 2003” was included in the health category due to the string “health” appearing in the title. Another issue is that phrases pertaining to multiple issues may be only identified in their primary context, as the project favors precision in the main topic of the text over recall.

Finally, we note that while new statements and votes are being added as they occur, the system is currently closed in terms of the candidates and issues supported. Adding a new candidate or politician currently requires identifying data sources for them, analyzing their structure, and writing code that can parse them. This system is quite vulnerable to small changes in the web page’s structure. If, in the future, all candidates support a standardized format to disseminate their views (e.g., using RSS feeds) and provide transcripts in a common layout, this would enable PSST to support very flexible queries involving a nearly unlimited number of politicians without any manual intervention. Adapting the system to a different country or set of candidates would then involve simply updating the issues, as some may be more or less relevant in the future and new issues may arise.

4. CONCLUSIONS AND FUTURE WORK

We presented PSST, a system for the analysis of political statements and votes, currently implemented for a predefined set of politicians and issues. Preliminary experiments support the validity of the approach. We plan to make improvements in the characterization of statements, integration of other data sources, and facilitation of expansion to include new candidates and issues. The project is located at: <http://cs.nyu.edu/~samantha/search/psst.html>

5. ACKNOWLEDGMENTS

The authors would like to thank Ernest Davis for helpful discussions.

6. REFERENCES

- [1] Extractor. www.extractorlive.com.
- [2] D. P. Kuhn. Bush’s top ten flip-flops: Cbsnews.com charts the opinion switches, part 1: George bush. www.cbsnews.com, September 2004.
- [3] D. P. Kuhn. Kerry’s top ten flip-flops: Cbsnews.com charts the opinion switches, part 2: John kerry. www.cbsnews.com, September 2004.
- [4] M. Sandalow. News analysis: Flip-flopping charge unsupported by facts. www.sfgate.com, September 2004.
- [5] P. Turney. Learning Algorithms for Keyphrase Extraction. *Information Retrieval*, 2(4):303–336, 2000.