Political Search Trends

Ingmar Weber
Yahoo! Research Barcelona
ingmar@yahoo-inc.com

Venkata Rama Kiran Garimella
Yahoo! Research Barcelona
gvrkiran@yahoo-inc.com

Erik Borra
University of Amsterdam
Amsterdam, Netherlands
borra@uva.nl

ABSTRACT

We present Political Search Trends, a browser based web search analysis tool that (i) assigns a political leaning to web search queries, (ii) detects trending political queries in a given week, and (iii) links search queries to fact-checked statements. In terms of methodology, it showcases the power of analyzing queries leading to clicks on selected, annotated web sites of interest.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services

Keywords

query log mining, political blogs, partisanship, fact-checking

1. WHAT IT DOES

Political Search Trends (PST)[1] analyzes anonymized web search queries submitted to Yahoo! that lead to clicks on political blogs. For a given week it shows trending *political* queries. E.g., the query "9 11 coloring book"[2] was found to be trending during Aug. 29-Sep. 4. As its main feature PST assigns a (fractional) political leaning to queries, such as "cost of obamacare" which is determined to be 99% right-leaning. Finally, it links queries to fact-checked statements on politifact.com such that "obama's aunt" is matched to `http://tinyurl.com/politifact-about-obamas-aunt`.

2. HOW IT WORKS

The starting point of PST is a set of blogs annotated as either left-leaning (387) or right-leaning (644). These labels are obtained by combining the work of Benkler and Shaw [1] with the Wonkosphere[3] blog directory. For the demo only queries leading to clicks on these blogs are considered. Obvious navigational queries are detected using string similarity and removed from consideration. Similarly, queries that within a given week only lead to clicks on one or two blogs are removed as they are generally not wide political issues. To improve the precision of detecting "political" queries, we compare the original number of web search results for a query

with the number of results when the search term (i) "political" or (ii) "celebrity" is added. In the first case, the difference must not be too large, and in the second it must not be too small. These checks remove a small number of queries such as "whitney houston's death", which lead to clicks on several borderline political blogs. Trends are detected using relative increases in search volume [2]. Queries are annotated as left- or right-leaning, on a 0-to-1 scale, according to the fraction of times a query leads to clicks on a particular leaning. To limit volume-induced leaning biases, we use *normalized* volumes where, conceptually, each leaning is given the same number of "votes" in a given week. To address the problem of sparsity, where a query with a small number of queries could prematurely be assigned an extreme leaning, we use appropriate smoothing. Queries are matched to fact-checked statements by first indexing both the statements, usually one sentence, and Politifact's short summary and then performing basic term-based matching, allowing incomplete fuzzy matches for long queries.

3. RELATED DEMOS

The tool most closely related to PST was Political Insights[4], also developed by us and now obsolete. The methodology for assigning a leaning to a query is the same for both tools [3]. However, PST includes the notions of time and "trending" and shows results for continuously updated fresh data. In terms of write-up, [3] focuses on the methodology and findings derived from the data, whereas here we focus on the demo. Apart from Political Insights, we are not aware of any "political" web search analysis tool. General search analysis tools[5] could be used to obtain historic volume information and a geographical distribution as well as, for Yahoo! Clues, demographic information about the users issuing the particular query. However, such tools do *not* present the political leaning of a query and the queries to explore need to be pre-defined by the user. PST on the other hand provides the user with a list of queries to explore for a given week or for a given search string. Media Cloud[6] uses partisan blogs but does not align them with search logs.

4. REFERENCES

[1] Y. Benkler and A. Shaw. A tale of two blogospheres: Discursive practices on the left and right, 2010. `http://tinyurl.com/Benker-and-Shaw`.

[2] I. Subasic and C. Castillo. The effects of query bursts on web search. In *Web Intelligence*, 2010.

[3] I. Weber, V. R. K. Garimella, and E. Borra. Mining web query logs to analyze political issues. In *Web Science*, 2012.

---

[1] `http://politicalsearchtrends.sandbox.yahoo.com`

[2] `http://abcn.ws/MyqKgU`

[3] `http://wonkosphere.com/directory.htm`

---

[4] Screenshots available at `http://bit.ly/KM9ieA`.

[5] `http://yhoo.it/lKSQ7t` or `http://bit.ly/tBzI`

[6] `http://www.mediacloud.org/`