

Blog Data Analytics Using Blogtrackers

Muhammad Nihal Hussain, Adewale Obadimu and Nitin Agarwal

University of Arkansas at Little Rock,

Little Rock AR 72204, USA

{mnhussain, amobadimu, nxagarwal}@ualr.edu

Abstract. Social media has grown to be the place for voicing one's opinions, sharing information, and shaping discourse. Individuals use social media as a platform to mobilize, coordinate, and conduct cyber campaigns ranging from awareness for diseases or disorders to deviant acts threatening democratic principles and institutions. Blogosphere has continued to rise and afford an effective medium for content framing. With no restriction on the number of characters, many use blogs to set narratives then use other social media channels like Twitter and Facebook to steer their audience to their blogs. Blog content is not structured and hard to collect than other social media channels. Blog monitoring and analysis could be of great use to sociologists, political scientists, communication researchers, journalists and information scientists to examine events. Toward this direction, we present Blogtrackers tool, which is designed to explore the blogosphere and gain insights on various events. Blogtrackers can help in identifying leading information actors, influential bloggers, popular and emerging trends, assess tones, sentiments and opinions, extract entities, and analyze their networks.

Keywords: Blog analysis, Blogtrackers, Blog monitoring, Social media.

1 Introduction

Blogging has become a popular channel as users can write without any restriction on topics that interest or affect them. Typically, other networking sites like Twitter and Facebook drive users to the blogs and empower discussions due to character limit. Blogosphere, defined as the network of blogs, is growing at an exponential rate. WordPress.com, which is one of the most popular blogging platforms, recorded activity of more than 87.6 million news posts added each month and over 409 million views per month in 2016 [1]. Analyzing blog data helps in understanding the pulse of a society, knowing what resonates with a community, and recognizing grievances of a group, among other reasons. Since there is no character limit in blogs unlike Twitter, blogs improve quality and inclusiveness of discourse and afford an effective platform to set narratives. Blogs also provide a convenient platform to develop situational awareness during a socio-political crisis or humanitarian crisis in a conflict-torn region or a disaster-struck area. In this demo, we introduce Blogtrackers, a tool that can be used for analyzing blogs to gain situation awareness.

2 Related Work

Due to a vast amount of publicly available data in the blogosphere [2], there is a need for creating usable tools for extracting actionable knowledge. Many blog tracking tools emerged in the past but these attempts have been largely discontinued. Some of the noteworthy tools include - BlogPulse, Blogdex, BlogScope, and Technorati. BlogPulse was developed by IntelliSeek. It was developed to provide search and analytic capabilities, automated web discovery for blogs, show the trends of information, track conversation and monitor the daily activity on weblogs. This tool was discontinued in 2012 [3]. Blogdex was another service that has been discontinued; it provided a resource for understanding hot-button issues in the blogosphere [4]. BlogScope was another blog tracking service developed as a research project in the department of computer science in university of Toronto, which provided blog analytics and visualizations but was shut down in early 2012 [5]. Technorati was originally launched as a blog index and search engine. It used a proprietary search and ranking algorithm to provide a rich directory of blogs sorted by content type and authority. However, it did not provide blog monitoring or analytical capabilities to the end users. Furthermore, Technorati stopped its blog indexing service May 2014 [6]. The service now offers advertising platform. Recognizing the need for a blog-monitoring tool in the research and practitioner community, we developed Blogtrackers (<http://blogtrackers.host.ualr.edu/>), which is discussed next.

3 Blogtrackers

Blogtrackers is a tool designed to explore the blogosphere and gain insights about events and how these events are perceived in the blogging community. Blogtrackers provides an analyst with means to develop situation awareness. Following are a few features and analytical capabilities of Blogtrackers:

1. Setup Tracker: A tracker is a collection of blogs selected by user for analysis. Setup Tracker page allows user to search for a topic of interest and select blogs to create a tracker for analysis. User can create and save any number of trackers.
2. Dashboard gives the overview of the selected tracker. It displays the number of blogs, bloggers, blog posts, total positive and negative sentiments. It also displays blog sites' hosting location and language distribution. Fig 2 shows the dashboard for the APAC blogs.
3. Posting Frequency can be utilized to identify any unusual patterns in blog postings. This aids in detecting real-time events that interested the blogging community. This feature also displays a list of active bloggers with number of posts. User can click on any data point on the graph to get a detailed list of the named-entities that were mentioned in blogs during that time-period.
4. Keyword Trends provides an overall trend of keywords of interest. It helps track changes in topics of interest in the blogging community. An analyst can correlate keyword trends with events to examine discussion topics and themes relating to that event. The analyst can select any data point on the trend line to view all the blogs. Fig 3 shows the keyword trends related to the APAC blogs
5. Sentiments and Tonality: displays the trend of positive and negative sentiments of blogs for any selected time-period (Fig. 1). This helps in understanding the effect an event has on the blogosphere. Additionally, data analyst can drill down by clicking on any point of interest and

view radar charts displaying tonality attributes such as personal concerns, time orientation, core drives, cognitive process.

6. Influence: This feature helps identify the influence a blogger or blog post has on the blogosphere. Blogtrackers finds the posts that are authoritative by assigning a score calculated using the iFinder model [7], [8]. This feature lists top 5 influential bloggers and displays a trend line to show the variation in bloggers' influence. Clicking on a point on the trend line allows a deeper dive into the data. This feature also provides capability to visually distinguish between influential and active bloggers. Further a user can explore the content themes of active-influential, inactive-influential, active-noninfluential, and inactive-noninfluential bloggers.

7. Additional Blog Info provides additional information about a blog. It gives a day-of-the-week average trend of a blog that helps in determining if the blog is a professional blog or a hobby blog. Also provided are monthly posting trend and sentiments for the past three years to determine the variation in activity and emotions. A list of URLs and domains mentioned in the blog is provided to know the source of information.

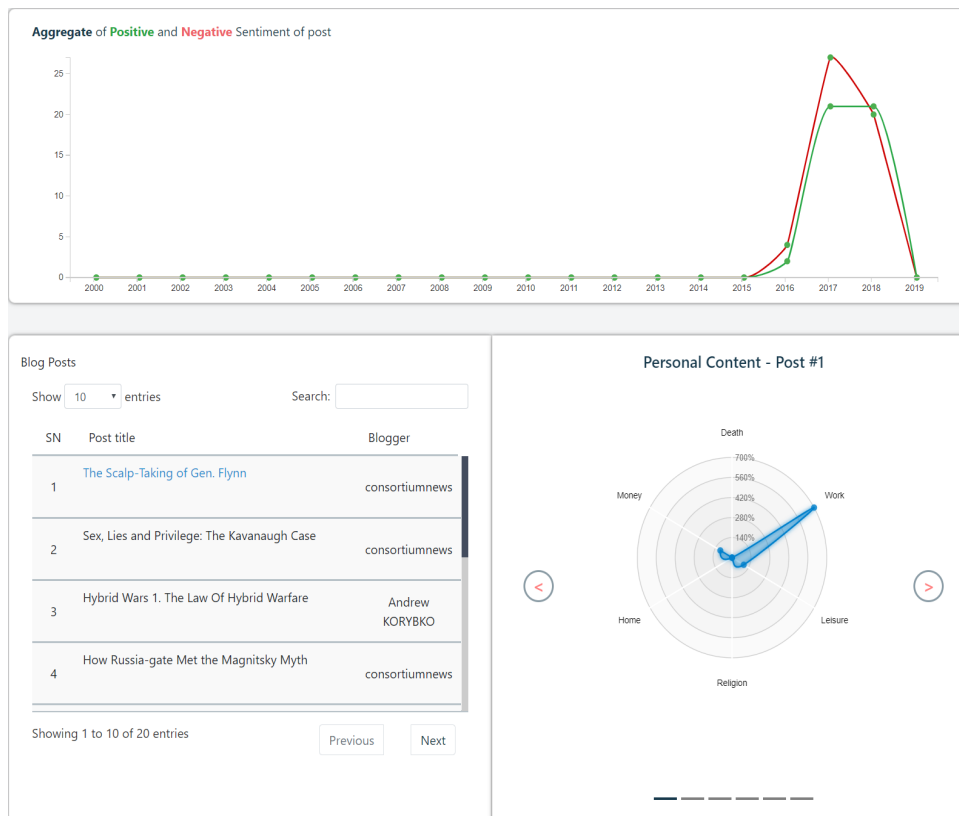


Fig. 1. Positive and negative sentiment analysis displaying tonality of blogs in "APAC" tracker.

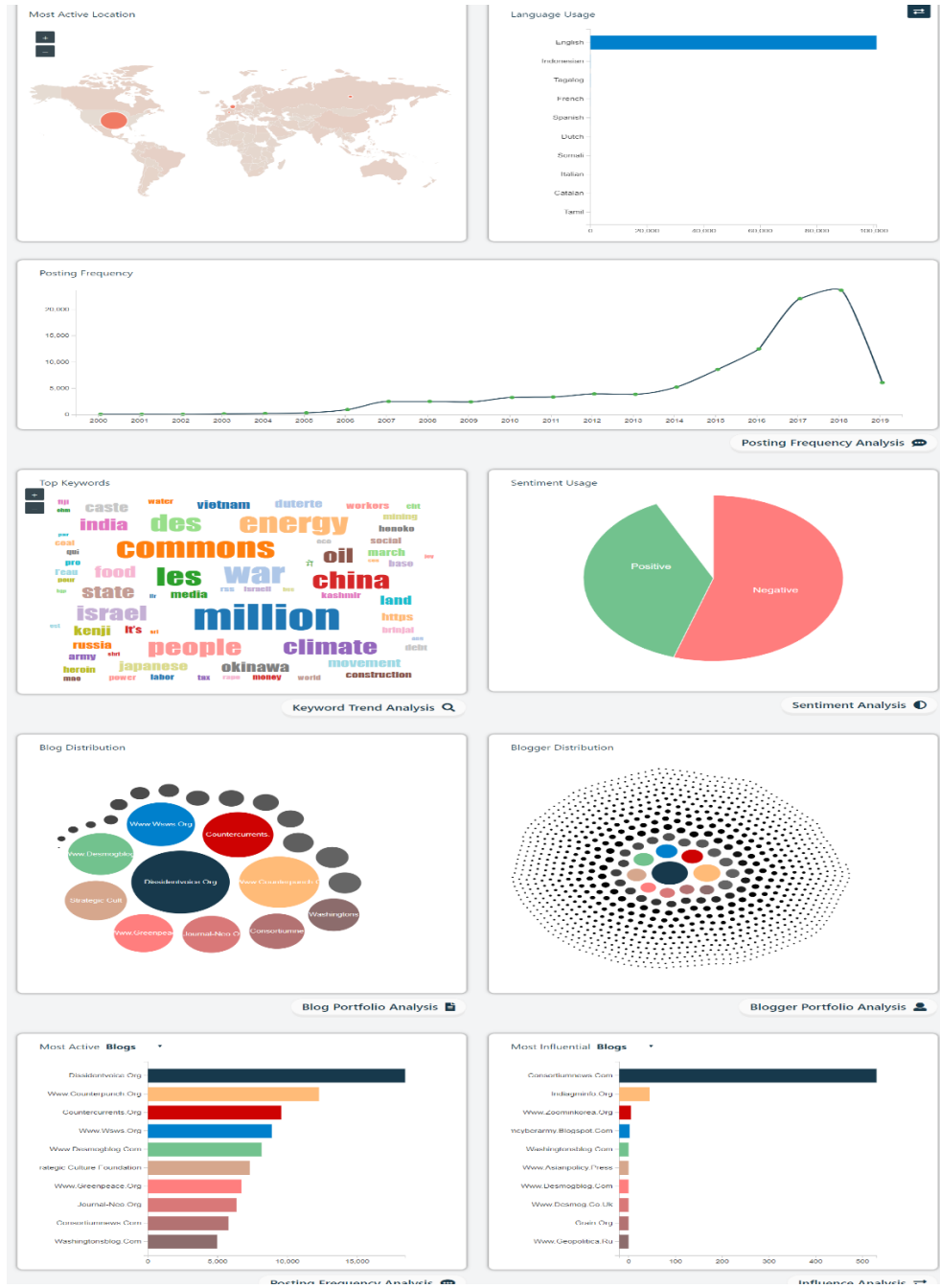


Fig 2. Dashboard for "APAC" tracker.

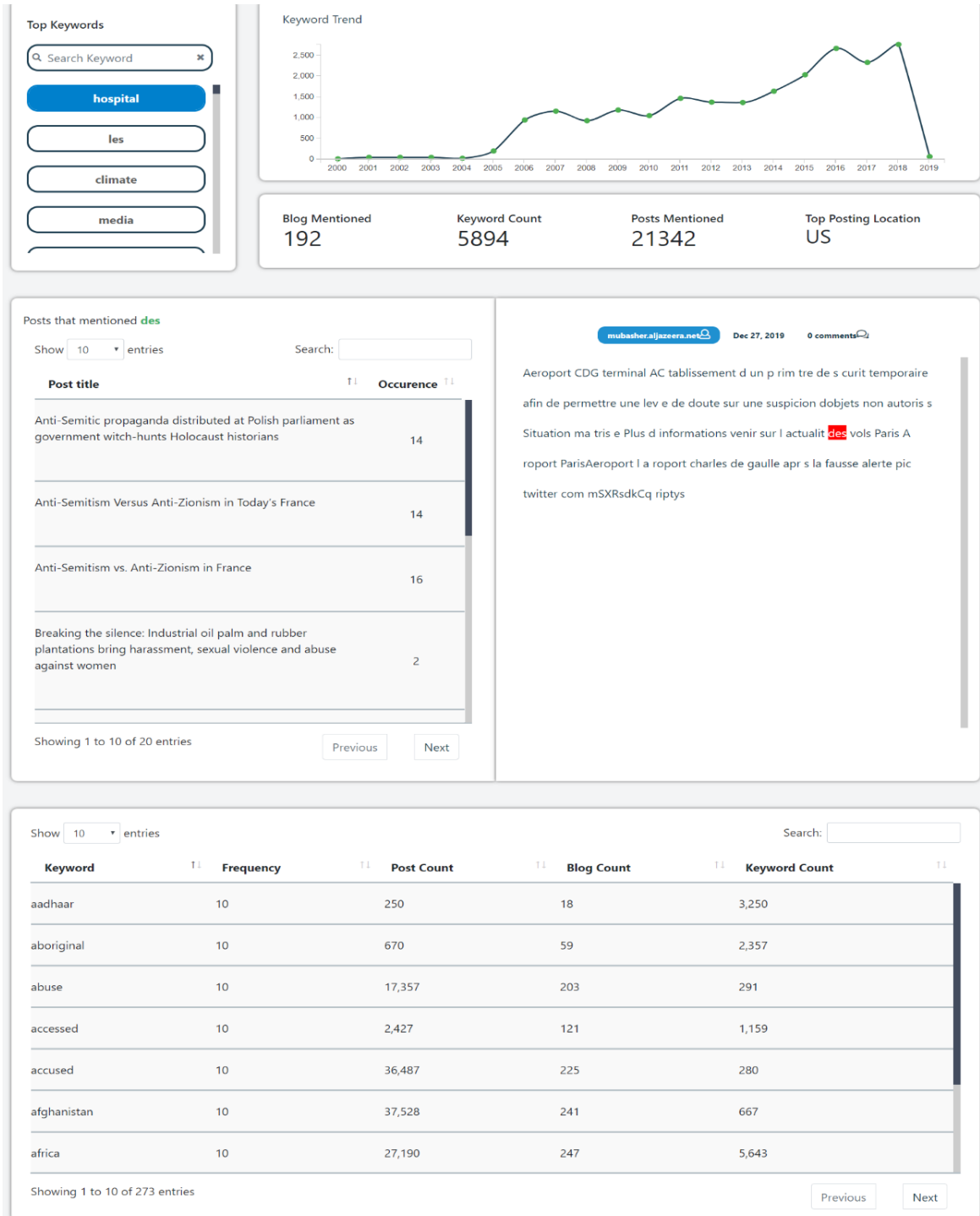


Fig. 3. Keyword Trends of "APAC" tracker

4 Conclusion and Future Work

In this paper, we explained the lack and need for blog tracking tools. We introduced the Blogtrackers tool. We are constantly improving Blogtrackers and adding new capabilities such as discourse analysis feature to study how narratives evolve over the course of an event. We understand that conversation travels from one channel to another and having a feature in Blogtrackers that enables cross-media analysis by building a network blogs, websites, media sharing sites like YouTube, and other social media websites like Facebook and Twitter would provide richer insights.

Acknowledgment

This research is funded by the U.S. Office of Naval Research (N000141010091, N000141410489, N0001415P1187, N000141612016, and N000141612412), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059) and the Jerry L. Maulden/Entergy Fund at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

- [1] “Stats — WordPress.com.” [Online]. Available: <https://wordpress.com/activity/>. [Accessed: 04-Apr-2017].
- [2] T. Furukawa, M. Ishizuka, Y. Matsuo, I. Ohmukai, K. Uchiyama, and others, “Analyzing reading behavior by blog mining,” in *Proceedings of the National Conference on Artificial Intelligence*, 2007, vol. 22, p. 1353.
- [3] “BlogPulse,” *Wikipedia*. 08-Mar-2017.
- [4] “Blogdex,” *Wikipedia*. 04-Nov-2016.
- [5] N. Bansal and N. Koudas, “Blogsphere: a system for online analysis of high volume text streams,” in *Proceedings of the 33rd international conference on Very large data bases*, 2007, pp. 1410–1413.
- [6] “Technorati—the World’s Largest Blog Directory—is Gone,” *Business 2 Community*. [Online]. Available: <http://www.business2community.com/social-media/technorati-worlds-largest-blog-directory-gone-0915716>. [Accessed: 04-Apr-2017].
- [7] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “Identifying the influential bloggers in a community,” in *Proceedings of the 2008 international conference on web search and data mining*, 2008, pp. 207–218.
- [8] N. Agarwal, H. Liu, L. Tang, and S. Y. Philip, “Modeling blogger influence in a community,” *Soc. Netw. Anal. Min.*, vol. 2, no. 2, pp. 139–162, 2012.