

Supporting Digital Epidemiology in Alberta via Twitter Tracking

Hamman Samuel, Osmar Zaiane, Patricia Martz

Abstract— In this ongoing work, we present a framework for tracking Twitter usage within the region of Alberta, Canada with the aim of ultimately supporting digital epidemiology tools. We have collected nearly 100,000 verified tweets over a period of 6 months. In addition, we have developed search and visualization tools for trend and geographical spread analysis.

I. INTRODUCTION

Public discourse on social media websites can provide valuable opportunities for tracking and predicting disease outbreaks. Searching for specific keywords within social media chatter can lead to a better understanding about health symptoms and risks being experienced [1]. In Canada, 25% of the population in the 18-64 age range uses Twitter, from which 23% are from Alberta, amounting to approximately 1.4 million users [2]. Digital epidemiology platforms such as the Alberta Real Time Syndromic Surveillance Network (ARTSSN) [3] currently use statistics from hospitals, and could benefit from additional real-time information streams such as social media discussions.

II. METHODOLOGY

We collect data from Twitter using its stream and search Application Programming Interfaces (APIs), with implementations in Python. The stream API allows retrieval of a small percentage of live tweets, while the search API gives access to a limited number of historical tweets. The stream API allows filtering of real-time tweets from within a bounded geographical area. To query the stream API for tweets made within Alberta, we approximate the region of Alberta as inscribed in a rectangular area. The search API also allows geographical filtering of tweets using a granularity parameter, which we set to Alberta as an administrative region. All retrieved tweets are stored in a MongoDB database. It should be noted that both APIs use internal heuristics to determine a tweet's location when the tweet's actual geographical coordinates are unavailable. The heuristics use the tweet author's profile location and time zone to predict the tweet's location.

III. RESULTS

From the period of July 2016 till January 2017, a total of 897,987 tweets were collected from the search and stream APIs. Out of these, 124,909 tweets had a longitude and

latitude, out of which 92,558 were verified to be in Alberta. We verified only those tweets that had coordinates recorded, using the Point-In-Polygon (PIP) method [4]. We also built search and visualization tools for keyword trend and geographic spread analysis, shown in Figure 1. The trend graph shows the frequency of health-related keywords over time, while the time map displays keywords on a map to provide an overview of geographical spread. In addition, a time slider can be controlled to view changes of geographical spread over time. For future work, we will focus on automatic identification of health keywords within tweets using the seven dimensions of wellness [5].

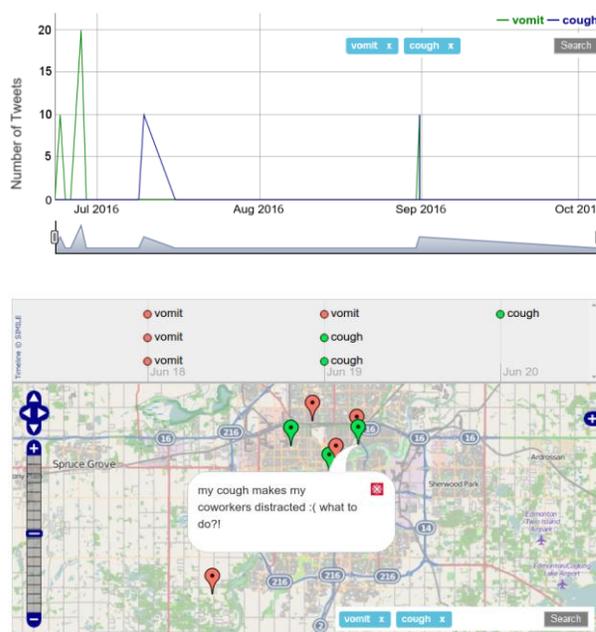


Figure 1. Trend Graph (Top) and Time Map (Bottom) Visualization Tools

REFERENCES

- [1] S. Brennan, A. Sadilek, H. Kautz. "Towards Understanding Global Spread of Disease from Everyday Interpersonal Interactions". In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), pp. 2783–2789, AAAI Press, 2013.
- [2] L. Bozinoff, "Instagram Tops in User Satisfaction," January 2015, Forum Research Inc.
- [3] P. Smetanin, R. K. Biel, D. Stiff, D. McNeil, L. Svenson, H. R. Usman, D. P. Meurer, J. Huang, V. Nardelli, C. Sikora et al., "An Early Warning Influenza Model using Alberta Real-Time Syndromic Data (ARTSSN)," Online Journal of Public Health Informatics (OJPHI), Vol. 7, No. 1, 2015.
- [4] M. Shimrat. "Algorithm 112: Position of Point Relative to Polygon". Communications of the ACM, 5(8):434, 1962.
- [5] L. T. Foster, C. P. Keller. British Columbia Atlas of Wellness. Western Geographical Press, Dept. of Geography, University of Victoria, 2007

*This research is funded by the Alberta Machine Intelligence Institute, Edmonton, Alberta, Canada

Hamman Samuel is a PhD Candidate with the Dept. of Computing Science, University of Alberta, Canada (E-mail: hwsamuel@ualberta.ca).

Dr. Osmar Zaiane is a McCalla & Killam Professor at the Dept. of Computing Science, University of Alberta (Email: zaiane@ualberta.ca).

Patricia Martz has a Master's degree from the School of Public Health, University of Alberta (Email: martz@ualberta.ca).