The Potential of Twitter for Early Warning and Outbreak Detection

Ed de Quincey, Patty Kostkova, Gawesh Jawaheer

City ehealth Research Centre, City University, London, UK

Abstract and Objectives

The use of user-generated content in Web 2.0 tools for predicting outbreaks has been seen as a great potential, however, the recent swine flu outbreak in April-May 2009 truly demonstrated the potential of these media for early warning systems. We have run a pilot study collecting tweets containing words related to influenza since May 2009 and collected over a million tweets until August 2009. An evaluation and data mining of this unique database is ongoing while the preliminary results are very promising.

Keywords:

Infection, Social networking, Outbreak detection, Web 2.0.

Introduction

Web 2.0 has generated a great interest recently as a possible media for early warning system for outbreak detection and epidemic intelligence (EI). Traditional systems such as GPHIN, Medisys and TTT are well established and used by ECDC and WHO on a daily bases, however, there has been recent interest in the ability to estimate flu activity via aggregating online search queries for keywords relating to flu and its symptoms by commercial companies like Google [1]. However, the search data remain proprietary and therefore not useful for research. The increase in user generated content on the web via social networking services such as Facebook and Twitter [2], however, provides researchers with a highly accessible view into people's online and offline activity.

Twitter Use Swine Outbreak Methods

Twitter, a micro-blogging service that allows people to post and read other users' 140 character messages currently has over 15 million unique users per month. Twitter allow third parties to search user messages using an open source API and return the text along with information from the poster's profile, such as their location, in a format that can be easily stored and analysed.

We searched and collected over 1 million tweets in the period from May until August 2009 and carry on collecting them on a minute bases to understand public concerns, keywords used and the profile of users who discuss these topics on the web [3].

Preliminary Results

We found over 1 million tweets reporting flu related illnesses and symptoms via Twitter in this period [4]. Most popular words in tweets were these (frequency in brackets):

flu (138, 260)
Swine (99, 179)
Have (13, 534)
Cases (13, 300)
H1N1 (9, 134)
Has (8, 010)
and others.

The actual sentence "I have swine flu" appeared 2, 888 times and "I have flu" 1,530 times. Further evaluation of the collected tweets, semantic relationship of keywords, geo-location of posters is underway and will be presented at the conference.

Discussion

The potential of Web 2.0 system for early warning systems and for better understanding public concerns about their health is enormous, however, further research is required to reveal the underlying principles and implement adequate integration with existing healthcare services.

References

[1] http://www.google.org/flutrends/

[2] http://www.crunchbase.com/company/twitter

[3] de Quincey E & Kostkova P. Using online social networking services like twitter to detect disease outbreak. Abstract in the Proceedings of the Health Protection Agency conference (HPA 2009), Warwick, UK, 14-16 September 2009

[4] de Quincey E, Kostkova P Early Warning and Outbreak Detection Using Social Networking Websites: the Potential of Twitter, Electronic Healthcare. 2nd International Conference, eHealth 2009, Intanbul, Turkey, September 2009. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, in press