ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Social Web mining and exploitation for serious applications: Technosocial Predictive Analytics and related technologies for public health, environmental and national security surveillance

Maged N. Kamel Boulos[a,*], *Antonio P. Sanfilippo*[b], *Courtney D. Corley*[b], *Steve Wheeler*[c]

[a] *Faculty of Health, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK*
[b] *Techno-Social Predictive Analytics Initiative, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, USA*
[c] *Faculty of Education, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK*

ARTICLE INFO

Article history:
Received 3 December 2009
Received in revised form
10 February 2010
Accepted 12 February 2010

Keywords:
Social Web
Web 2.0
Disease surveillance
Technosocial Predictive Analytics

ABSTRACT

This paper explores Technosocial Predictive Analytics (TPA) and related methods for Web "data mining" where users' posts and queries are garnered from Social Web ("Web 2.0") tools such as blogs, micro-blogging and social networking sites to form coherent representations of real-time health events. The paper includes a brief introduction to commonly used Social Web tools such as mashups and aggregators, and maps their exponential growth as an open architecture of participation for the masses and an emerging way to gain insight about people's collective health status of whole populations. Several health related tool examples are described and demonstrated as practical means through which health professionals might create clear location specific pictures of epidemiological data such as flu outbreaks.

© 2010 Elsevier Ireland Ltd. All rights reserved.

1. Background

The rapid development and growing popularity of the Social Web, also commonly known as "Web 2.0", and its applications, coupled with the increasing availability of mobile Internet-enabled devices (such as netbooks and Apple iPhone), has ensured that the Internet is now a truly integral part of everyday life. The Social Web is a particularly dynamic medium that captures the "pulse" of the society in real time. Blogs, micro-blogs such as Twitter (http://twitter.com/), and social networking sites such as Facebook (http://www.facebook.com/), enable people to publish their personal stories, opinions, product reviews and a great deal more in real time. They can also share geo-tagged alerts and reports about their current location and allow others to follow their whereabouts, all in real time, using location-aware social services such as Microsoft Vine (http://www.vine.net/). Such applications have greater power when coupled to Social Web 'mashup' services and aggregators. Such tools weave data from disparate sources into a new, compound data source or service [1,2]. A useful example of the latter is HealthMap, the global disease alert map (http://www.healthmap.org/en).

* *Corresponding author at*: Room 05, 3 Portland Villas, Faculty of Health, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK. Tel.: +44 0 1752 586530; fax: +44 0 7053 487881.
E-mail addresses: mnkamelboulos@plymouth.ac.uk (M.N. Kamel Boulos), antonio.sanfilippo@pnl.gov (A.P. Sanfilippo), Courtney.Corley@pnl.gov (C.D. Corley), S.Wheeler@plymouth.ac.uk (S. Wheeler).
0169-2607/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2010.02.007

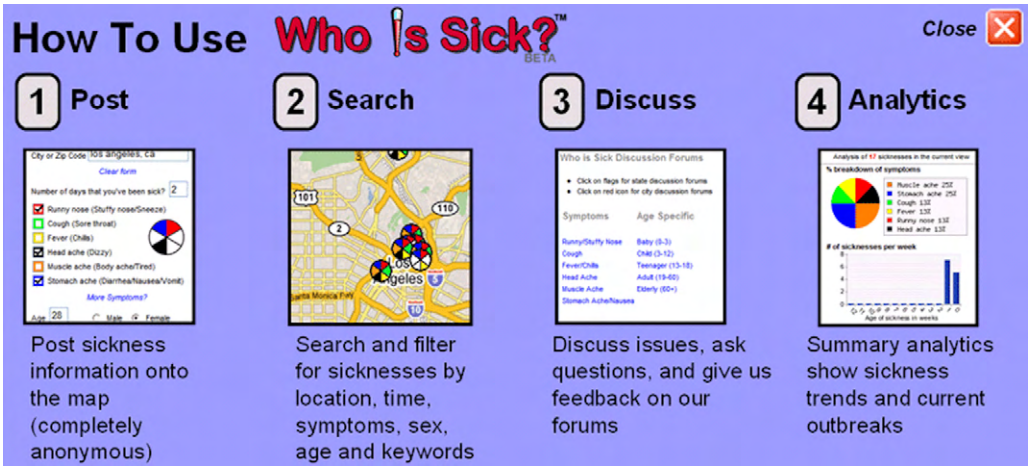

Fig. 1 – 'How To Use Who Is Sick?' (Screenshot from http://whoissick.org/sickness/).

The dynamic nature and continuously updated features of the Social Web make it a fertile environment for intelligence gathering in a variety of disciplines, enabling users to tap into the 'wisdom of the crowds'. Further, the Social Web is useful for gleaning the collective 'impression of the masses' regarding current matters, events and products [3]. In fact, it is not uncommon these days for example, to see shoppers (including those who do not make their purchases through the Internet) consulting online user reviews and ratings on sites such as Amazon.com about products they are planning to buy before making their purchase decisions.

This paper will briefly review a number of emerging technologies and tools, including Technosocial Predictive Analytics pioneered at Pacific Northwest National Laboratory, USA (http://predictiveanalytics.pnl.gov/), that can be used to exploit these inherent Social Web features in real or near-real time and harness them for public health, environmental and national security surveillance purposes.

2. A brief introduction to Technosocial Predictive Analytics and some related technologies and tool examples

2.1. *Who Is Sick: the power of user-generated maps*

'Who Is Sick' (http://whoissick.org/sickness/, Fig. 1) was launched in 2006 as a location-aware service for tracking and monitoring sickness, with the aim of providing current and local sickness information *by the public to the public*. People can anonymously post information about their own illness and use a familiar Google Maps interface to search and filter illness by symptoms, sex, age, and location. The maps act as a crude syndromic surveillance tool in real time. A person feeling unwell can instantly check what sicknesses are circulating within their own locale. Health-conscious individuals conduct similar scans and take any necessary preventive measures in preparation. People traveling to another area of the country or abroad can use the service to learn if there are any sicknesses going around that they need to be aware of. One should be warned however, that information quality is always dependent upon the accuracy of user-reported details and on sufficient numbers of people (who are actually sick) using the tool to report their symptoms in a timely manner.

2.2. *We Feel Fine: differentiating between informative and affective text on the Social Web*

Created in 2006, 'We Feel Fine' (http://www.wefeelfine.org/) is a Web-based applet and API (Application Programming Interface—http://wefeelfine.org/api.html) that can be used to assess the fluctuating effects of real world events and factors such as the Stock Index or the weather on people's feelings and emotions, and ascertain a crude idea about the 'general mood' of populations at different dates and locations around the world [4].

At the heart of 'We Feel Fine' is a data crawler that automatically scours the Web every 10 min, harvesting data in the form of human feelings from a large number of blogs and social pages such as those hosted by MySpace (http://www.myspace.com/) and Blogger (http://www.blogger.com/). 'We Feel Fine scans' blog posts for occurrences of the phrases "I feel" and "I am feeling", and once any of these is found, the system goes backward to the beginning of the sentence and forward to the end of the sentence, and then saves the full sentence in a database. Saved sentences are next scanned against a manually compiled list of about 5000 pre-identified "feelings" (adjectives and some adverbs). If a valid feeling is found, the corresponding sentence is considered to represent one person who feels that way. 'We Feel Fine' also attempts to extract the username of the post's author from the URL of his/her blog post (e.g., 'steve-wheeler' in http://steve-wheeler.blogspot.com/) and uses it to automatically traverse the blog hosting site to locate that user profile page. From the profile page, it is often possible to extract the age, gender, country, state, and city of residence of the blog owner. 'We Feel Fine' can also retrieve the local weather conditions for the blog owner's city at the time the post was written. This process results in about 15,000 to 20,000 saved feelings (with associated data) per day. For display purposes, the top 200 feelings were manually assigned colours that loosely correspond to the tone of the feeling, e.g., happy

Download English Version:

https://daneshyari.com/en/article/468012

Download Persian Version:

https://daneshyari.com/article/468012

Daneshyari.com