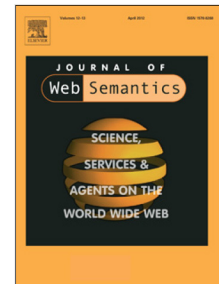


Accepted Manuscript

A framework for real-time semantic social media analysis

Diana Maynard, Ian Roberts, Mark A. Greenwood, Dominic Rout,
Kalina Bontcheva



PII: S1570-8268(17)30024-0

DOI: <http://dx.doi.org/10.1016/j.websem.2017.05.002>

Reference: WEBSEM 434

To appear in: *Web Semantics: Science, Services and Agents on the World Wide Web*

Received date: 14 October 2016

Revised date: 13 March 2017

Accepted date: 5 May 2017

Please cite this article as: D. Maynard, et al., A framework for real-time semantic social media analysis, *Web Semantics: Science, Services and Agents on the World Wide Web* (2017), <http://dx.doi.org/10.1016/j.websem.2017.05.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Framework for Real-time Semantic Social Media Analysis

Diana Maynard, Ian Roberts, Mark A. Greenwood, Dominic Rout, Kalina Bontcheva

*University of Sheffield
Department of Computer Science
Regent Court, 211 Portobello
S1 4DP, Sheffield, UK
d.maynard@sheffield.ac.uk*

Abstract

This paper presents a framework for collecting and analysing large volume social media content. The real-time analytics framework comprises semantic annotation, Linked Open Data, semantic search, and dynamic result aggregation components. In addition, exploratory search and sense-making are supported through information visualisation interfaces, such as co-occurrence matrices, term clouds, treemaps, and choropleths. There is also an interactive semantic search interface (Prospector), where users can save, refine, and analyse the results of semantic search queries over time. Practical use of the framework is exemplified through three case studies: a general scenario analysing tweets from UK politicians and the public's response to them in the run up to the 2015 UK general election, an investigation of attitudes towards climate change expressed by these politicians and the public, via their engagement with environmental topics, and an analysis of public tweets leading up to the UK's referendum on leaving the EU (Brexit) in 2016. The paper also presents a brief evaluation and discussion of some of the key text analysis components, which are specifically adapted to the domain and task, and demonstrate scalability and efficiency of our toolkit in the case studies.

Keywords: Natural Language Processing, semantic search, social media analysis, Linked Open Data, semantic annotation, sentiment analysis

1. Introduction

Social media is the largest collection of information about society that we have ever had, providing an incredibly rich source of behavioural evidence. However, understanding and using it in a meaningful way is often still a major problem. Gleaning the right information can be tricky because analytics tools either do not provide the right kinds of interpretation, or are simply not accurate, aggregated, enriched or easily interpretable.¹ In the recent 2015 UK elections, for example, numerous analytics tools attempted to understand the attitudes of the public towards the various parties and to predict the outcome of the election, but mostly with quite poor results as they did not take into account many subtle nuances. There are many reasons for this, which are not appropriate to discuss here, but one reason is that investigating people's values, and their opinions on

specific topics such as the economy, rather than their opinions on particular parties as a whole, seems to give better insight.² Furthermore, simple sentiment analysis tools that look at people's opinions [1] often do not deal well with nuances such as sarcasm, nor the fact that people tend to express their sentiment about very specific events rather than about a party overall, which may have subtle differences. We therefore need much more sophisticated forms of analysis in order to understand properly what people are saying.

Social media content is dynamic, reflecting the societal and sentimental fluctuations of the authors. User activities on social networking sites are often triggered by popular or specific events and related entities (e.g. sports events, celebrations, crises, news articles) and topics (e.g. global warming, terrorism or immigration).

The unique nature of social media data is precisely what makes it also so challenging [2]. It is fast-growing,

¹<http://simplymeasured.com/blog/2015/03/09/5-problems-with-how-marketers-use-social-analytics/>

²<http://www.theguardian.com/politics/2015/may/14/why-did-the-election-pollsters-get-it-so-wrong>

Download English Version:

<https://daneshyari.com/en/article/6950518>

Download Persian Version:

<https://daneshyari.com/article/6950518>

[Daneshyari.com](https://daneshyari.com)