ARTICLE IN PRESS

Information Processing and Management xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Onyx: A Linked Data approach to emotion representation

J. Fernando Sánchez-Rada *, Carlos A. Iglesias

Intelligent Systems Group, Departamento de Ingeniería de Sistemas Telemáticos, Universidad Politécnica de Madrid (UPM), Spain

ARTICLE INFO

Article history:
Received 21 May 2014
Received in revised form 26 March 2015
Accepted 31 March 2015
Available online xxxx

Keywords:
Emotion analysis
Ontology
Provenance
Linked data
EmotionML
Lexical resources

ABSTRACT

Extracting opinions and emotions from text is becoming increasingly important, especially since the advent of micro-blogging and social networking. Opinion mining is particularly popular and now gathers many public services, datasets and lexical resources. Unfortunately, there are few available lexical and semantic resources for emotion recognition that could foster the development of new emotion aware services and applications. The diversity of theories of emotion and the absence of a common vocabulary are two of the main barriers to the development of such resources. This situation motivated the creation of Onyx, a semantic vocabulary of emotions with a focus on lexical resources and emotion analysis services. It follows a linguistic Linked Data approach, it is aligned with the Provenance Ontology, and it has been integrated with the Lexicon Model for Ontologies (lemon), a popular RDF model for representing lexical entries. This approach also means a new and interesting way to work with different theories of emotion. As part of this work, Onyx has been aligned with EmotionML and WordNet-Affect.

1. Introduction

With the rise of social media, more and more users are sharing their opinions and emotions online (Pang & Lee, 2008). The increasing volume of information and number of users are drawing the attention of researchers and companies alike, which seek not only academical results but also profitable applications such as brand monitoring. As a result, many tools and services have been created to enrich or make sense out of human generated content. Unfortunately, they are isolated data silos or tools that use very different annotation schemata. Even worse, the scarce available resources are also suffering from the heterogeneity of formats and models of emotion, making it hard to combine different resources.

Linked Data can change this situation, with its lingua franca for data representation as well as a set of tools to process and share such information. Plenty of services have already embraced the Linked Data concepts and are providing tools to interconnect the previously closed silos of information (Tummarello, Delbru, & Oren, 2007). In fact, the Linked Data approach has proven useful for fields like Opinion Mining. Some schemata offer semantic representation of opinions (Westerski, Iglesias, & Tapia, 2011), allowing richer processing and interoperability.

Emotions have a crucial role in our lives, and even change the way we communicate (Pang & Lee, 2008). They can be passed on just like any other kind of information, in what some authors call emotional contagion (Barsade, 2002). That is a phenomenon that is clearly visible in social networks (Kramer, Guillory, & Hancock, 2014). Public APIs make it relatively easy to study the social networks and their information flow. For this very reason Social Network Analysis is an active field (Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee, 2007), with Emotion Mining as one of its components. On the other

* Corresponding author.
E-mail addresses: jfernando@dit.upm.es (J.F. Sánchez-Rada), cif@gsi.dit.upm.es (C.A. Iglesias).

http://dx.doi.org/10.1016/j.ipm.2015.03.007
0306-4573/

Please cite this article in press as: Sánchez-Rada, J. F., & Iglesias, C. A. Onyx: A Linked Data approach to emotion representation. *Information Processing and Management* (2015), http://dx.doi.org/10.1016/j.ipm.2015.03.007

hand, the growing popularity of services like micro-blogging will inevitably lead to services that exchange and use emotion in their interactions. Some social sites are already using emotions natively, giving their users the chance to share emotions or use them in queries. A noteworthy example is Facebook, which recently updated the way its users can share personal statuses.

Nevertheless, the impact of emotion analysis goes well beyond social networks. For instance, there are a variety of systems whose only human–machine communication is purely text-based. Such systems can use the emotive information to change their behavior and responses (Pang & Lee, 2008).

Furthermore, there are many sources that can be used for sentiment analysis beyond pure text, including video and audio. Multimodal analysis, or making use of several of these sources, is an active field that gathers experts from different disciplines. A unified schema and appropriate tooling would open up new possibilities in this field.

Lastly, combining subjective information from emotion analysis with facts already published as Linked Data could enable a wide array of new services. This would require a widely accepted Linked Data representation for emotions, which does not exist yet. In this paper we present Onyx, a new vocabulary that aims to bridge this gap and allow for interoperable tools and resources. We also provide a set of example applications, additional vocabularies to use existing models, and multilingual resources that use Onyx to annotate emotion.

The rest of this paper is structured as follows. Section 2 introduces the technologies that Onyx is based upon, as well as the challenges related to emotion analysis and creating a standard model for emotions, including a succinct overview of the formats currently in use. Section 3 covers the Onyx ontology in detail, including some vocabularies or models of emotions, and several examples. Section 4 presents the results of our evaluation of the Ontology, focusing on the coverage of current formats like Emotion Markup Language (EmotionML). Finally, Section 5 completes this paper with conclusions and future work.

2. Enabling technologies

2.1. Models for emotions and emotion analysis

To work with emotions and reason about them, we first need to have a solid understanding and model of emotions. This, however, turns out to be a rather complex task. It is comprised of two main components: modeling (including categorization) and representation.

Confusingly, the terms opinion, sentiment, emotion, feeling and affect are commonly used interchangeably. Throughout this article we follow the terminology by Cambria, Schuller, Xia, and Havasi (2013) where opinion mining and sentiment analysis are focused on polarity detection and emotion recognition, respectively.[1]

There are also several models for emotions, ranging from the most simplistic and ancient that come from Chinese philosophers to the most modern theories that refine and expand older models (Ekman, 1999; Prinz, 2004). The literature on the topic is vast, and it is out of the scope of this paper to reproduce it. The recent work by Cambria, Livingstone, and Hussain (2012a) contains a comprehensive state of the art on the topic, as well as an introduction to a novel model, *the Hourglass of emotions*, inspired by Plutchik's studies (Plutchik, 1980). Plutchik's model is a model of categories that has been extensively used (Borth, Chen, Ji, & Chang, 2013; Cambria, Havasi, & Hussain, 2012b; Mohammad & Turney, 2013) in the area of emotion analysis and affective computing, relating all the different emotions to each other in what is called the wheel of emotions.

All the existing motions are mainly divided in two groups: discrete and dimensional models. In discrete models, emotions belong to one of a predefined set of categories, which varies from model to model. In dimensional models, an emotion is represented by the value in different axes or dimensions. A third category, mixed models, merges both views.

Other models are more general and model affects, including emotions as a subset. One of them is the work done by Strapparava and Valitutti in WordNet-Affect (Strapparava et al., 2004), an affective lexicon on top of WordNet. WordNet-Affect comprises more than 300 affective labels linked by concept–superconcept relationships, many of which are considered emotions. What makes this categorization interesting is that it effectively provides a taxonomy of emotions. It both gives information about relationships between emotions and makes it possible to decide the level of granularity of the emotions expressed. Section 3.2.1 discusses how we formalized this taxonomy using SKOS, and converted it into an Onyx vocabulary.

Despite all, there does not seem to be a universally accepted model for emotions (Schröder, Pirker, & Lamolle, 2006). This complicates the task of representing emotions. In a discussion regarding EmotionML, Schroder et al. pose that *given the fact that even emotion theorists have very diverse definitions of what an emotion is, and that very different representations have been proposed in different research strand, any attempt to propose a standard way of representing emotions for technological contexts seems doomed to fail* (Schröder et al., 2007). Instead, they claim that the markup should offer users choice of representation, including the option to specify the affective state that is being labeled, different emotional dimensions and appraisal scales. The level of intensity completes the definition of an affect in their proposal.

[1] A more detailed terminology discussion can be found in Munezero, Montero, Sutinen, and Pajunen (2014).