TweetSpector: Entity-based retrieval of Tweets

Surender Reddy Yerva, Zoltán Miklós, Flavia Grosan, Alexandru Tandrau, Karl Aberer
Swiss Federal Institute of Technology (EPFL)
Lausanne, Switzerland
{surenderreddy.yerva,zoltan.miklos,flavia.grosan,alexandru.tandrau,karl.aberer}@epfl.ch

Categories and Subject Descriptors

H.3.1 [**Information Systems Applications**]: Content Analysis and Indexing; H.3.5 [**Information Systems Applications**]: Online Information Services

Keywords

Entity, Disambiguation, Profiles, Twitter

1. EXTENDED ABSTRACT

People readily express their opinions about the various products, companies, TV shows etc., on Twitter[1]. These tweet messages are thus a rich source of information that can be exploited to understand the sentiments about the concerned products or services. Retrieving the tweets related to given entities is however a challenging task as their names are often (deliberately) ambiguous, e.g. Apple, Blackberry, Friends, etc. Nevertheless, identifying the relevant entities is an essential first step to develop reliable sentiment analysis techniques that is not considered in existing systems, for example TweetFeel[2], TwitterSentiment[3].

While there is a number of techniques for identifying named entities in unstructured text, they are often not directly applicable in this case, as tweet messages are very short (maximal 140 characters). This demonstrator introduces TweetSpector, a tool that addresses this retrieval task and enables to link tweet messages to a given entity. Our retrieval methods rely on classification techniques that exploit our concise descriptions of entity-relevant information, also called entity profiles.

The demonstrator presents the following features of TweetSpector:

-*Entity Profile Creation:* TweetSpector supports automatic profile creation, where we apply named-entity recognition, NLTK, wordnet and Web data extraction techniques to construct profiles for an entity, given a relevant Webpage. TweetSpector also enables manual profile construction, where users can construct arbitrary entity profiles, as well as manual and automatic updates for initially constructed profiles (thus the profiles are dynamic). The profiles can also be visualized using Word Clouds.

---

[1] http://www.twitter.com

[2] http://www.tweetfeel.com

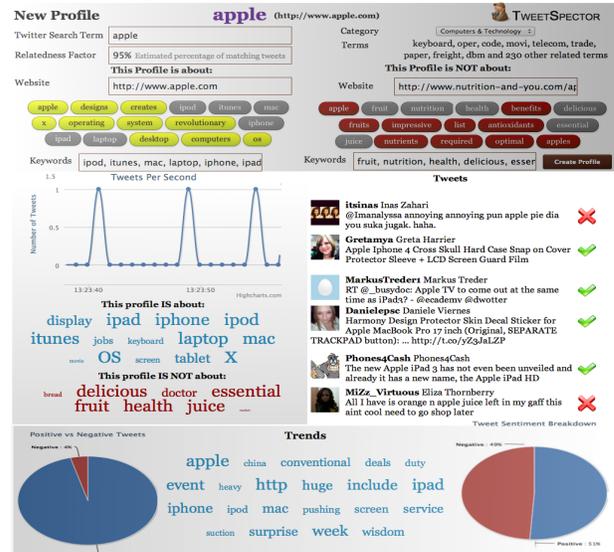[3] http://twittersentiment.appspot.com

Figure 1: TweetSpector: Various Features

-*Tweet Classification:* TweetSpector displays in real-time the classification results (see Figure 1). For example, a stream of tweets is displayed and it is indicated whether or not the messages shall be related to the company Apple Inc.. The classification techniques are widely extended versions of our earlier work [1].

-*User Feedback:* The users can indicate whether the proposed classification is correct or not. This feedback is taken into account by the algorithms. TweetSpector can also take human input through crowdsourcing (through an interface to Amazon Mechanical Turk).

-*Dashboard:* TweetSpector can display performance metrics and statistical information on a dashboard related to the entity.

2. ACKNOWLEDGEMENTS

3. REFERENCES

[1] Surender Reddy Yerva, Zoltán Miklós, and Karl Aberer. Entity-based Classification of Twitter Messages. *International Journal of Computer Science & Applications*, 9(1):88–115, 2012.