On the enrichment of a RDF Repository of City Points of Interest based on Social Data

Zied Sellami, Gianluca Quercini, Chantal Reynaud-Delaître

On the Enrichment of a RDF Repository of City Points of Interest based on Social Data

Zied Sellami
LRI, Université Paris-Sud XI
91405 Orsay, France
zied.sellami@lri.fr

Gianluca Quercini
Supélec, E3S
91900 Gif-sur-Yvette, France
gianluca.quercini@supelec.fr

Chantal Reynaud-Delaître
LRI, Univ. Paris-Sud XI,CNRS
91405 Orsay, France
chantal.reynaud@lri.fr

ABSTRACT

Points of interest (POIs) in a city are specific locations that present some significance to people; examples include *restaurants*, *museums*, *hotels*, *theatres* and landmarks, just to name a few. Due to their role in our social and economic life, POIs have been increasingly gaining the attention of location-based applications, such as online maps and social networking sites. While it is relatively easy to find on the Web basic information about a POI, such as its geographic location, telephone number and opening hours, it is more challenging to have a deeper knowledge as to what other people say about it. What if a person wants to know all the restaurants in Paris that serve good seafood and provide a kind service? Typically, the answer to this question has to be looked for on websites that let people leave comments and opinions on POIs, a time-consuming manual task that few are willing to do. This search would be better supported by search engines if information mined from opinions were available in a structured form, such as RDF. In this position paper, we describe a general approach to enrich an existing RDF repository about POIs with data obtained from social networking sites.

1. INTRODUCTION

Points of interest (POIs) are geographic entities that present some significance to people because they play a specific role in a city. Examples include places where we eat (*restaurants*), sleep (*hotels*), have a good time (*pubs/nightclubs*) or engage in cultural activities (*museums/theatres*). As they are an important part of our social and economic life, POIs are the focus of many location-based applications, such as online maps, mobile apps and social networks [14].

While it is relatively easy to find on the Web basic information about a POI, such as its geographic location, telephone number and opening hours, it is more challenging to have a deeper knowledge as to what other people say about it. What if a person wants to know all the restaurants in Paris that serve good seafood and provide a kind service? Websites exist, such as *TripAdvisor.com*, that allow people to write reviews about POIs, which, however, are in a highly unstructured form and therefore not processable easily by a search engine. As a result, answers to complex queries that take into account people opinions on POIs cannot be determined automatically and typically people need to sift through a lot of web pages to find any partial response.

In this paper we argue that the wealth of social data on POIs can be used to automatically enrich existing RDF repositories; more specifically, we can use the user comments and reviews to get a deeper knowledge on POIs, namely what is good or bad about it, and represent it in a structured and machine-readable form. This goal entails a number of very interesting research problems:

- Given a reference to a POI in an existing RDF repository, find all pages across social networking sites (SNSs) that describe the POI.
- Analyse the user comments and opinions found in the SNS pages to understand whether the general opinion on the POI is positive and what is positive or negative about it.
- Represent the outcome of the opinion mining in a structured form, such as RDF.

In this paper, we focus on the first two problems, while we leave the third as an interesting direction for future work.

The problem of matching a POI across multiple SNSs is related to entity reconciliation in databases, which consists of determining whether two distinct table records refer to the same real-world entity. Although many approaches to entity reconciliation have been proposed, to the best of our knowledge none has been evaluated on POIs. Moreover, in our case we need to look for the occurrence of P in a social network based on the values of very few facets. The fact that very often the location of P is known helps the reconciliation, but it does not make the problem straightforward. Indeed, often is the location approximate, which prevents us from resorting to an exact matching of location information. Therefore, we propose a reconciliation method that uses a measure that computes a similarity score of any pair of POI occurrences based on the known facets.

Once the pages relative to a POI are found, we can extract from them useful information. Typically, SNSs, such as *Foursquare* or *Yelp*, provide APIs that make easy the extraction of metadata about a POI, including reviews. In this paper we analyse NLP techniques to rate a POI based on the reviews found on SNSs pages and also to understand what is good or bad about it.

Our work is motivated by specific application needs. In the context of the research project *DataBridges: Data Integration for Digital Cities*, a 2011-2012 activity within the "Digital Cities of the Future" action line of the *EIT ICT Labs* KIC, we developed a *faceted browser* over a RDF repository of POIs obtained by automatically extracting data from *Google Fusion Tables* (GFT) [12, 13, 17]. In our application POIs are organized according to different attributes or *facets* (e.g., geographic location, telephone number, category), which are used to explore the POIs by applying multiple filters. As a result, it is possible to visualize in few clicks information on Italian restaurants in Paris by filtering by POI type (restaurant), category (Italian cuisine) and location (Paris). However, while we retrieved a large number of POIs from GFT, we could only identify values for few facets, which severely limits the browsing experience provided by our application. For instance, GFT tables usually provide the geographic location (address and geographic coordinates) of restaurants and, less frequently, the URL of their website and telephone numbers; rarely could we find information as to the cuisine type, dress code, menu price or reviews. Yet, this information is extremely valuable.

In the remainder of the paper we review existing research work that relate to ours (Section 2), we describe the similarity measures that we use to match a POI across different SNSs (Section 4) and we detail our approach to inferring information on a POI based on user opinions (Section 5). A preliminary evaluation is presented in Section 6 to validate our approach; we conclude our presentation in Section 7.

2. RELATED WORK

The work introduced in this paper entails different interesting research problems, of which we identify three:

- Entity reconciliation. Given the reference to a POI in a RDF repository and a page on a social network site, does the page describe the given POI in the repository?
- Sentiment analysis. Given the reviews of users that can be found on the page of a social network, what can be inferred from it? Is the general opinion on the POI positive or negative? What are the specific positive and negative aspects? Can we infer some recommendations from them?
- Representation of opinions in a structured form. How can we represent the outcome of a sentiment analysis in a structured form such as RDF? Ontologies for opinions?

Although many approaches to entity reconciliation have been proposed, to the best of our knowledge none has been evaluated on POIs. The most related approach is the one proposed by Scheffler et al. [16], to match references to POIs across multiple social networks based on the similarity of their names and the pages describing them. To reduce the number of comparisons, their approach uses geographic coordinates to narrow down the search space to nearest POIs.

Lot of approaches have been proposed for sentiment analysis and we refer the reader to [8] for an extensive and up-to-date review. Most of the proposed approaches deal with the determination of the polarity (positive, negative or neutral) of users comments. Pak and Paroubek [10] and Read [15] propose a classifier that identifies whether a tweet expresses a positive or negative mood based on the presence of emoticons, such as ":-)" and ":-(". Tools are available on the Web to detect the polarity of tweets such as *Sentiment140* [1] and *SMM* [2]. Sentiment analysis relies on linguistic approaches to detect positive or negative expressions in a text. Some approaches identify the polarity of a sentence based on the adjectives [9]; others rely on verbs and adjectives combined [4]. For instance, in the sentence "This camera is great", the adjective "great" indicates a positive opinion about a "camera" and in the sentence "I do not like this photo", the verbal expression "not like" indicates a negative opinion about a "photo". In this paper, we present how we used this kind of approach to extract positive/negative aspects about a POI.

As for the representation of opinions in a structured form, there is little research at the moment. [18] proposes an ontology for opinion mining called *Marl*. However, the ontology seems to be limited and not complete. Although in this paper we do not investigate further this aspect, we believe that it might be an interesting direction for future work. [3] proposes a tool for enriching and visualising POIs called *TOPICA*. Additional information like POI category, POI keywords (city, country, etc.) or images are extracted from DBPedia and showed in a world map. However, no sentiment analysis process based on users reviews is proposed.

3. PROBLEM DEFINITION

Let R be a RDF repository of data on POIs in several cities across the world. In this paper we refer to R as the repository that we obtained by automatically extracting and annotating data from *Google Fusion Tables* (GFT) while developing a faceted browser over RDF data [13, 17]. R is described by an ontology that we manually created to represent the properties of POIs that are used in our browser. More specifically, each POI is assigned a set of *facets*, or *properties*, of which the most important are *name*, *geographic location*, *type* (e.g., "restaurant", "museum", "hotel") and *category* ("italian restaurant", "archaeological museum", "four-star hotel"). These properties apply to any POI in R. Other properties, such as *price*, *dress code*, *owner name*, *phone number*, *website*, are specific to some types of POIs (e.g., restaurants, museums, hotels).

Since R has been created through an automatic extraction and annotation procedure, it is not unlikely to contain incomplete or partial or sometimes inaccurate information. This might depend on either the procedure itself or the source of the data, in this specific case the GFT tables. In particular, we observed that GFT tables contain a fair amount of data on POIs, which is not surprising given that they are directly contributed by people over the Internet and that POIs play such an important role on people lives. However, only few facets are usually reported, which generally include the name and the address, less frequently category and type and rarely price, telephone number and website [13]. As a result, R contains references to thousands POIs each having values for only a limited set of facets.

This motivates the work of our paper, which aims at enriching R by assigning more values to the facets of each POI, based on information available on social networking sites (SNSs). SNSs are an extraordinary source of data that are likely to be:

[1] http://www.sentiment140.com
[2] http://smm.streamcrab.com

1. *Up-to-date*, because people use SNSs on a daily basis;

2. *Complete*, especially regarding POIs, which are the focus of people's social interests.

3. *Accurate*, because many people double-check them constantly.

Note that the same arguments apply to other data sources, such as *Wikipedia*, which provides knowledge across several domains, including POIs; however, in Wikipedia some types of POIs, including hotels and restaurants, are underrepresented, which is not the case of SNSs such as *Foursquare* [3] and *Yelp* [4], which we use in this paper. Moreover, SNSs provide additional information that cannot be found in Wikipedia, such as people reviews and comments that we exploit to extract some useful insights about a POI, namely positive and negative aspects of it.

One of the major challenge we need to face is the need of linking any POI p of R to the corresponding web page $WP(p)$ of a SNS that contains information about p. Typically, $WP(p)$ contains information on p, such as the name or the address, that can be easily extracted, either by using an API provided by the SNS itself, which is the case of *Yelp* and *Foursquare*, or by crawling the page, which usually complies with a fixed template, as all pages within the same SNS have the same structure and appearance. For the sake of convenience, $WP(p).property$ (respectively, $p.property$) denotes a property of p that is found in the web page $WP(p)$ (respectively, in the repository R). Finding the web page that has information on p requires a comparison between the properties of p that are available in the repository R and those that are found in the web pages of the SNS. More specifically, if $p.name = WP(x).name$ and $p.address = WP(x).address$, where the symbol $=$ indicates an exact matching between strings, then it is fair to conclude that $WP(x) = WP(p)$. However, exact string matching works only in the ideal case where R and the SNS represent the values of the properties of the POI in the same way, which is often not the case. For instance, "The Louvre", "Louvre Museum", "Le Louvre" and "Musée du Louvre" are different names that refer to the same entity. When it comes to addresses, the situation is even worse. Addresses in fact are rarely represented in a standard and formatted way. Most of the addresses that we found in GFT tables are either *partial*, containing only a reference to the street name and, possibly, the city name, or *missing*, in which case they are usually represented with geographic coordinates. Unfortunately, even the geographic coordinates did not prove to be accurate in most cases. As a result, we need to resort to a measure that computes the *similarity*, rather than the *equality*, between the properties that are in R and those that are found in the SNS web pages.

4. MATCHING POIS ACROSS SOCIAL NETWORKS

Given a reference to a POI p in our repository R, we want to find the web page $WP(p)$ in a SNS that has information on p. We use the search engine of the SNS and we select only the top first candidate set C of web pages result which contains the most relevant results. For each web page $WP(x) \in C$ we use a measure that determines the similarity between $WP(x).name$ and $p.name$, as follows:

$$sim(p.name, WP(x).name) =$$

$$\frac{Levenshtein(p.name, WP(x).name) + Jaccard(p.name, WP(x).name)}{2}$$

Before applying the similarity measure, the values of $p.name$ and $WP(x).name$ are normalized by using the Porter stemmer [11] and removing stopwords. Our measure combines two well-known similarity coefficients, namely the *Levenshtein distance* [7] and the *Jaccard index* [6]. The novelty is the combined use of both, which is based on the observation that the two coefficients complement each other. In fact, the *Levenshtein distance* between two phrases is equal to the number of characters that need to be edited to change one phrase into the other; as a result, if two phrases are composed of exactly the same words, arranged in a different order, their *Levenshtein distance* turns out to be low. In order to balance this, we use the *Jaccard index*, which boosts the similarity score of two phrases that share most words even in a different order. Note that POIs are often referenced with names that comply with this rule: "Louvre Museum" and "Museum of Louvre" are just two examples.

We then determine that $WP(x)$ is the Web page that has information on the POI p if and only if the following is true:

$$sim(p.name, WP(x).name) \geq \delta_1 \vee$$

$$(sim(p.name, WP(x).name) \geq \delta_2 \wedge distance(p, WP(x)) \leq dist_{max})$$

where $\delta_1 > \delta_2$ and $distance(p, WP(x))$ is the distance between p and the POI described in the Web page $WP(x)$. The rationale of this formula is the following. If $p.name$ and $WP(x).name$ are *very* similar, which is indicated by a value of their similarity higher than the given threshold δ_1, then $WP(x)$ is considered to be the web page corresponding to p. Otherwise, the geographic information of p and $WP(x)$ are used to determine their distance, i.e. if the two POIs are close from a geographical point of view. If $p.name$ and $WP(x).name$ are still acceptably similar, i.e., their similarity is higher than a fixed threshold δ_2, and p and the POI described by $WP(x)$ are relatively close to each other, then $WP(x)$ is considered to be the page describing p.

As for the distance between p and the POI described by the web page $WP(x)$, we use the geographic coordinates; these might be directly available or need to be inferred from the address of the POI. To this extent, we use the Google Geocoding API, which can translate an (possibly, partial and ill-formatted) address into geographic coordinates. In Section 6 we discuss in greater detail the values of the thresholds δ_1, δ_2 and $dist_{max}$ that we use in our formula.

5. OPINION MINING

Once a link is established between a POI p in R and a web page $WP(p)$ in a SNS, we can use the information provided by the SNS to enrich R, namely by adding values to properties of P, such as the telephone number or the website. This can be done straightforwardly, by using the API provided by the SNS, if any, or extracting that information directly from $WP(p)$; recall that usually the pages of a SNS have the same graphic appearance, meaning that they are based on a fixed template, which makes the extraction of information across pages very easy.

In this section we go a step further and we propose a method to analyse the opinions that people leave on the pages of SNSs to extract relevant information about POIs, namely positive and negative aspects of it. This relates to opinion mining, which is a well-established problem in the social network community; however, most of the existing approaches to opinion mining can determine whether an opinion is positive or negative but do not specify what is positive or negative.

We go through the following two steps:

[3] http://foursquare.com/
[4] http://yelp.com/

1. We use a NLP technique to extract from a comment phrases that indicate a quality of a POI and/or a feeling of a person about the POI. We also determine if the phrases have a positive or negative connotation.
2. We use the extracted phrases to enrich R.

The two steps are detailed in the next subsections.

5.1 Phrase Extraction

Our approach is based on linguistic techniques. More precisely, we define two categories of linguistic expressions, including adjectives and/or verbs, to identify expressions of sentiments:

- Expressions describing POI or objects in a POI.
- Expressions indicating personal sentiments or advices of a person about a POI.

The first category of expressions indicate people appreciations about a POI or the quality of objects in a POI. We identified at least two patterns that belong to this category (the star occurring after an element indicates that the element is optional in the pattern):

1. **Pattern 1** : **(NOT)* (ADVERB)* ADJECTIVE OBJECT** (e.g., *Great food*, *not interesting place*);
2. **Pattern 2** : **OBJECT BE (ADVERB)* ADJECTIVE** (e.g., *sandwich is good*, *restaurant is nice*).

The second category of expressions is used by people to express their feelings on a POI and, possibly, some recommendations. We identified three patterns belonging to this category:

1. **Pattern 3** : **IT IS (ADVERB)* ADJECTIVE** (e.g., for instance *it's interesting*, *it's nice*);
2. **Pattern 4** : **I (NOT)* FEEL OR SUGGEST OBJECT** (e.g., *I like this place*, *I advice you to test this hotel*);
3. **Pattern 5** : **I FEEL (ADVERB)* ADJECTIVE** (e.g., *I feel happy*, *I feel very hungry.*

Each pattern is used as a regular expression to extract phrases from a comment that might indicate either a quality of a POI, or a recommendation or an opinion. We are then left with determining whether the extracted phrases indicate a positive or a negative aspect of a POI.

To this extent, we create four lexicons containing positive/negative adjectives and positive/negative verbs by processing with the POS tagger *Treetagger* [5] a corpus of positive and negative words[6] [5]. We identified 1467 positives adjectives, 1609 negatives adjectives, 421 positives verbs and 1243 negatives verbs.

5.2 Repository Enrichment

At this step, we have a collection of phrases extracted from the comments that people left on $WP(p)$ of a SNS to give their opinion on p. In order to add the information obtained from the comments to our repository R, we first obtain a general evaluation on p based on our collection of phrases. The general evaluation is composed of a *grade* and an *appreciation*. The grade is a value between -10 and 10 which is computed by Equation 1.

$$notation = \left(\frac{|Positive_{phrase}|}{|All_{phrase}|} - \frac{|Negative_{phrase}|}{|All_{phrase}|}\right) \times 10 \quad (1)$$

where $Positive_{phrase}$, $Negative_{phrase}$ and All_{phrase} denote respectively the set of positive, negative phrases and all phrases identified with the technique described in Subsection 5.1.

Each grade is associated with an appreciation: *Very Bad*, *Bad*, *Medium*, *undetermined*, *Fairly, Good* and *Very Good*. As a result, a POI is considered to be *Very Bad* if its grade is below -6.6, *Very Good* if its grade is above 6.6 and *undetermined* if its grade is 0.

Next, we add to R three categories of information about a POI:

1. **General assessment**. This kind of information is identified with **Pattern 1** and **Pattern 2** when the object of a phrase is the term *place*, the name of the POI or its type. Examples are *beautiful place*, *prestigious museum*, *cool place*. This information is also identified with **Pattern 3** and **Pattern 5** which give the sentiment of a reviewer about the POI (e.g., *I feel happy*).
2. **Tips**. This kind of information is identified with **Pattern 4**, when the phrase refers to a suggestion. Examples are *not use pyramid entrance*, *I advise you to visit basement.*
3. **Specific ideas**: This kind of information is identified with **Pattern 1** and **Pattern 2** when the object of the phrase is not the term *Place*, the name of the POI or its type. If so, the phrase refers to a specific aspect that a person likes or not. Examples are *beautiful art*, *incredible artwork*, *horrible food.*

For the moment, no method is used to select only the most important information, which can be done by applying statistical measures such as TF-IDF . This is reserved for future work.

6. EVALUATION

In this section we present the evaluation of our similarity approach and our opinion mining approach.

6.1 Evaluation of the Similarity Formula

To evaluate the relevance of our similarity formula, we selected 600 POIs from the RDF repository and we manually identified the corresponding pages on *Foursquare* (to have a ground truth). Then, we applied our similarity formula to each POI to determine the corresponding page of *Foursquare* and we evaluated the determinations against the ground truth by using precision (P), recall (R) and f-measure (F), defined as:

$$P = \frac{|C|}{|D|} \quad R = \frac{|C|}{|A|} \quad F = 2 \cdot \frac{P \cdot R}{P + R}$$

where C is the set of POIs for which the correct page is determined, D is the set of POIs for which a page (either correct or wrong) is determined and A is the set of all POIs. We compared the results against the Levenshtein distance and the Jaccard similarity. Since we aim at enriching a repository with **correct information**, we mostly aim at **improving precision** while having acceptably high values for recall. The results are showed in Table 1.

Formula	Precision	Recall	F-measure
Levenshtein	83.90%	68.01%	75.12%
Jaccard	85.56%	65.87%	74.44%
Our similarity formula	**86.08%**	**65.76%**	**74.56%**

Table 1: Comparison of our similarity formula against the Levenshtein distance and the Jaccard similarity

Table 1 shows that our similarity approach obtained the best precision. Levenshtein and Jaccard indeed have their own limitations, but when used in combination they are overcome. In fact, the similarity of two strings is measured at character-level by the Levenshtein distance and at word-level by Jaccard. As a result, the two strings *Restaurante Hotel Baltum* and *Hotel Baltum*, which refer to the same POI, are considered as dissimilar by the Levenshtein distance and similar by Jaccard; on the other side, the two strings *Musée le Louvre* and *Louvre Museum* are considered as dissimilar by Jaccard and similar by the Levenshtein distance. While combining both measures, we obtain the correct result in both cases.

[5]http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/
[6]http://www.cs.uic.edu/ liub/FBS/sentiment-analysis.html

6.2 Evaluation of the Opinion Mining Approach

To evaluate our opinion mining approach we analysed 40 comments on the *Louvre Museum* and the *Eiffel Tour* left by users on *Yelp*.

106 positive expressions and 17 negative expressions were identified from the *Louvre Museum* users reviews. Examples of positive aspects of the *Louvre* determined by our approach is that the place is "fantastic" and "amazing". Examples of negative aspects are that the place is *massive* and *unconfortable*. The result is detailed in Table 2.

Expressions categories	Positive	Negative
Global Ideas about the POI	16	3
Tips given by reviewers	0	1
Specific Ideas about the POI	80	12

Table 2: Detail of the opinion mining analysis of the Louvre Museum

105 positive expressions and 8 negatives expressions were identified from the *Eiffel Tour* reviews. By using our opinion mining approach we retrieve that reviewers suggest people *to go to the top of the tour*. We also find out that the view is nice, great and panoramic and that one negative aspect is the *illegal Eiffel tower souvenirs* and the *crazy line*. The result is detailed in the table 3.

Expressions categories	Positive	Negative
Global Ideas about the POI	8	0
Tips given by reviewers	1	0
Specific Ideas about the POI	79	8

Table 3: Detail of the opinion mining analysis of the Eiffel Tour

7. CONCLUSION

In this paper we presented an approach to enrich a RDF repository of POIs based on social data. In particular, we focused on two research problems that arise in this context: the identification of a POI across multiple social networking sites and the extraction of information from the reviews that users leave on these sites. For the first problem we investigate a similarity measure that combines two well-known similarity metrics, such as the Levenshtein distance and the Jaccard similarity. We see that their combination allows to overcome their limitations. However, the similarity measure as it is has room to be improved. Indeed, for now we use only few facets to compare the POIs and it would be interesting to try different combinations of facets and similarity measures.

Our opinion mining approach shows encouraging results, but needs to be improved. We are studying a learning approach based on SentiWordNet[1] to complete the list of positive and negative adjectives and verbs. In fact, in the experiment of our opinion mining approach, some expressions identified with patterns, do not have a polarity value because the adjective and/or the verb in the expression does not exist in our lexicon lists. Also, we want to include in our opinion mining approach the use of adverbs to precise the polarity of the expression [2]. For instance, the sentence *the food is not too bad*, has a positive sentiment, because of the presence of the adverb "too".

8. REFERENCES

[1] Baccianella, S., Esuli, A., and Sebastiani, F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)* (Valletta, Malta, May 2010), European Language Resources Association (ELRA).

[2] Benamara, F., Irit, S., Cesarano, C., Federico, N., and Reforgiato, D. Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone. *In Proc of Int Conf on Weblogs and Social Media* (2007).

[3] Cano, A. E., Burel, G., Dadzie, A.-S., and Ciravegna, F. TOPICA: A Tool for Visualising Emerging Semantics of POIs based on Social Awareness. In *10th International Semantic Web Conference (ISWC 2011)* (2011).

[4] Chesley, P., Vincent, B., Xu, L., and Srihari, R. Using verbs and adjectives to automatically classify blog sentiment. In *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)* (2006), pp. 27–29.

[5] Hu, M., and Liu, B. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2004), KDD '04, ACM, pp. 168–177.

[6] Jaccard, P. *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*. Bulletin de la Société vaudoise des sciences naturelles. Impr. Corbaz, 1901.

[7] Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. Tech. rep., 1966.

[8] Liu, B. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies 5*, 1 (2012), 1–167.

[9] Moghaddam, S., and Popowich, F. Opinion polarity identification through adjectives. *CoRR abs/1011.4623* (2010).

[10] Pak, A., and Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (Valletta, Malta, May 2010), European Language Resources Association (ELRA).

[11] Porter, M. An algorithm for suffix stripping. *Program 14*, 3 (1980), 130–137.

[12] Quercini, G., and Reynaud, C. Entity Discovery and Annotation in Tables. In *Accepted for publication at EDBT2013* (2013).

[13] Quercini, G., Setz, J., Sonntag, D., and Reynaud, C. Facetted browsing of extracted fusion tables data for digital cities. In *Proceedings of the Web of Linked Entities Workshop in conjunction with the 11th International Semantic Web Conference (ISWC 2012)* (2012), pp. 94 – 105.

[14] Rae, A., Murdock, V., Popescu, A., and Bouchard, H. Mining the Web for Points of Interest. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2012), SIGIR '12, ACM, pp. 711–720.

[15] Read, J. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop* (Stroudsburg, PA, USA, 2005), ACLstudent '05, Association for Computational Linguistics, pp. 43–48.

[16] Scheffler, T., Schirru, R., and Lehmann, P. Matching Points of Interest from Different Social Networking Sites. In *35th Annual German Conference on Artificial Intelligence* (2012), Ki 2012, pp. 245–248.

[17] Setz, J., Quercini, G., Sonntag, D., and Reynaud, C. Facetted search on extracted fusion tables data for digital cities. In *35th Annual German Conference on Artificial Intelligence (Demo paper)* (2012).

[18] Westerski, A., Iglesias, C. A., and Rico, F. T. Linked Opinions: Describing Sentiments on the Structured Web of Data. In *4th International Workshop Social Data on the Web (SDoW2011)* (2011).