

Tour-pedia: a web application for the analysis and visualization of opinions for tourism domain

Andrea Marchetti*, Maurizio Tesconi*, Stefano Abbate*, Angelica Lo Duca*, Andrea D'Errico*, Francesca Frontini**, Monica Monachini**

* IIT – CNR Pisa Italy {name.surname}@iit.cnr.it

** ILC – CNR Pisa Italy {name.surname}@ilc.cnr.it

Abstract

We present Tour-pedia an interactive web application that extracts opinions from reviews of accommodations from different sources available on-line. Polarity markers display on a map the different opinions. This tool is intended to help business operators to manage reputation on-line.

Keywords: Visualization tools, opinion mining, NLP on social media, tourism reviews

1. Introduction

OpeNER (Open Polarity Enhanced Name Entity Recognition) is a project funded under the 7th Framework Program of the European Commission. It has a duration of 2 years. It started officially in July 2012 and will terminate in June 2014. Partners collaborating in the project are SMEs and research institutions from Italy, Spain and the Netherlands.

The main objective of OpeNER is to provide a set of ready-to-use modules for the processing of natural language. More specifically, OpeNER focuses on building linguistic pipelines in six languages (English, Spanish, German, French, Italian, Dutch) that enable the identification and disambiguation of named entities and the analysis of sentiment in opinionated texts. The foreseen application is to extract the attitude of customers regarding given topics (such as hotels and accommodations) in online reviews.

This implies on the one hand the analysis of large quantities of data (such as reviews and feedback left by users on dedicated platforms and on social media), and on the other hand the aggregation of such data for the benefit of professionals within the tourism industry, so that they can assess the effectiveness of their marketing strategies and better direct their efforts.

The named entities that are found in reviews can belong to the general domain, such as persons, locations, dates, figures, etc. but also to the more specific tourism domain, such as accommodations, attractions, point of interests. In the first case the entities can be recognized and disambiguated using generic tools such as Wikipedia-DBpedia (this is what happens in the out-of-the-box version of DBpedia Spotlight); in the second case a specific resource was required to integrate the others. This resource, that we present here, is called Tour-Pedia, and is conceived as an encyclopedia of tourism.

2. Tour-pedia

The resource has been initialized using the data gathered within the project focusing on three areas: Amsterdam, Tuscany, Spain, representing respectively a metropolis, a region and a nation. The data concern both available accommodations and reviews on such accommodations,

containing potential opinions on such entities. The final objective is to extend the project on the whole of Europe. Figure 1 shows the architecture of Tour-Pedia together with the steps that are required for its building and accessing.

As is shown by the figure, the data are gathered both from the Social Media such as Facebook, FourSquare, GooglePlaces, Booking, TripAdvisor and from archives published as Linked Data. Great attention is given to problems of licensing, by collecting only information that is labeled as open, such as hotel name, address etc., or taking contact with the providers directly when dealing with non-open information.

A particularly complex problem was the integration of the various sources, since the information gathered from the social media has proved highly unreliable (with mistakes on the geographic coordinates and imprecisions in the addresses) and often duplicated. For this reason we have contacted organizations for tourism operating in the targeted areas in order to obtain the official archives to be used as pivots in the matching operations.

The thus constructed archive (Named Entity Repository) presently contains information on accommodations (Venues: hotels, hostels, camping, bed&breakfasts, ...) as well as a collection of reviews related to them. The reviews have been collected by crawling on social media and specialized web platforms, and have been annotated using the OpeNER pipelines (Analyzed Reviews)

Table 1 and 2 show figures on the current state of the archive, which can be queried by using a restFul Web API.

Source	Amsterdam	Tuscany	Spain
Booking	669	TBD	TBD
Fondazione Sistema Toscana	-	19.140	-
Google Places	952	12.587	44.706
Foursquare	775	4.528	14.471
Facebook	1226	3.414	9.908

Table 1 – Accommodations

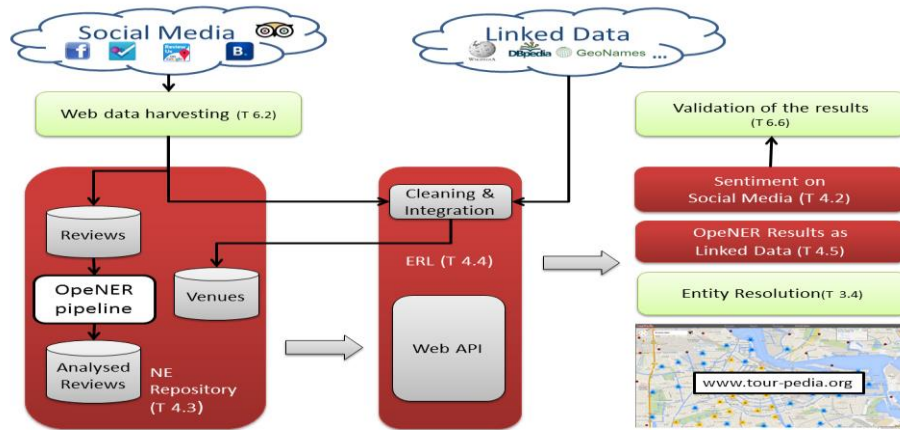


Figure 1 – Building Tour-pedia.

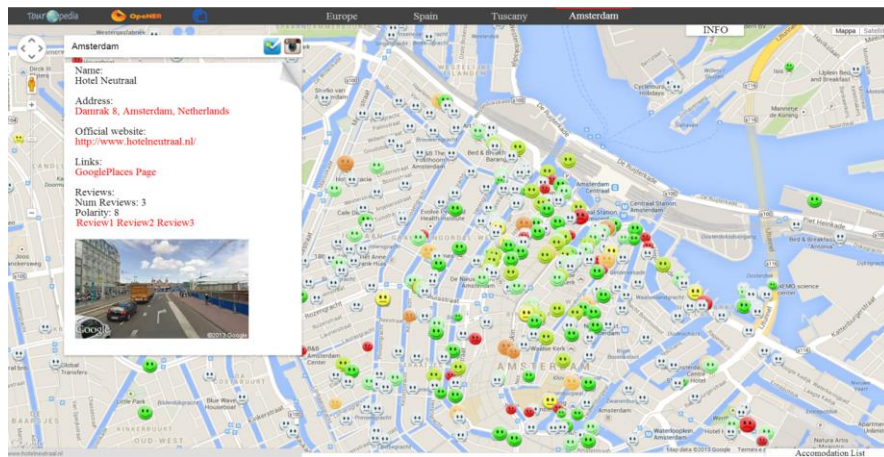


Figure 2 – Tour-pedia GUI.

Language	Number of Analyzed Reviews
Dutch	646
English	14.819
French	3.912
German	1.206
Italian	7.928
Spanish	36.004

Table 2 - Reviews

3. GUI

Besides the API access, a specific graphical interface was designed for accessing in a simple and immediate way.

We conducted an analysis of existing GUIs in similar projects, and the best example appears to be the interface of Google Maps for its minimal, and intuitive look and feel. We therefore chose to emulate the style of navigation of Google Maps, by potentiating and enhancing its fundamental characteristics:

- a map occupying the whole page
- a simple menu place over the map
- a search bar inserted inside the map itself

Moreover, in order to concentrate the attention of the user on the map (the focal point of the interface) all info-windows appear over the map, without subtracting too much space from it.

Accommodations appear on the map as smileys: the color and mood of the icon is determined by aggregating the sentiment extracted from the reviews for that accommodation. An overall number of positive sentiment will produce a green smiley, red is for overall negative sentiment, and intermediate colors express ranges. White locations have no reviews available for evaluation.

4. Conclusions and future works

TourPedia is a reference archive for tourism, containing accommodations and reviews from three European locations: Amsterdam, Tuscany, Spain. It was created within the framework of the OpeNER project in order to produce a reference archive to disambiguate entities in the tourism domain. In view of making its use possible also outside the project, an access interface was produced and will be made available also as a SPARQL end-point according to the Linked Data principles.

The next steps will be, alongside with the export in Linked Data, also the addition of attractions and points of interests alongside with the accommodations to the archive