TOPICA: A Tool for Visualising Emerging Semantics of POIs based on Social Awareness Streams

A. Elizabeth Cano[1], Gregoire Burel[2], Aba-Sah Dadzie[1], and Fabio Ciravegna[1]

[1] OAK Group, Dept. of Computer Science, The University of Sheffield, UK
[2] Knowledge Media Institute, The Open University, Milton Keynes, UK
{a.cano,a.dadzie,f.ciravegna}@dcs.shef.ac.uk,g.burel@open.ac.uk

Abstract. *Topica* is an application that enriches the Social Web with semantic data, to enable collective perception of Points of Interest (POIs), which are human constructs that describe information about locations (e.g., restaurants, attractions, cities). *Topica* provides an extra layer of information, compared to existing applications for browsing POIs, by modelling hidden characteristics of POIs, by: (1) generating a Linked Data representation of the collective perception of a POI; (2) enhancing the POI representation by mashing up services that enrich the POI's related entities; and (3) providing a visual representation of the POIs adapted to suit user- and context-sensitive filters. *Topica* identifies topics relevant to a POI by extracting DBpedia categories from *entities* (e.g., People, Places) and *keywords* (e.g., Crete, Bonn) obtained from social awareness streams related to the POIs.

Key words: Linked Data, Semantic Web, Point of Interest, Social Awareness Streams, citizen-sensing, social data mining, emerging semantics

1 Introduction

The availability of multiple social media platforms, such as Facebook[3] and Tripadvisor[4], usable from desktops and mobile phones, has altered how the ordinary end user engages with their social context and environment. Particularly, *social awareness streams*, collections of semi-public, natural-language messages distinguished by their brevity, have resulted in new communication patterns characterised by high social connectivity and active trending of topics [2]. Recent studies in user profiling have proposed the use of social awareness streams to model users' interests, activities and behaviour [1, 4, 6].

However, little work exists that uses social awareness streams to model location entities. Stevie, for instance [3], allows users to share and browse temporal information about points of interest (POIs) – events – on a map; the POIs are enriched using concepts extracted from a set of ontologies and the location broadcast by end users' GPS (Global Positioning System). Tintarev et al. [5] demonstrate the added benefit in personalising recommendations of popular POIs for tourists.

Scope for extending work in this area exists. The W3C POI Working Group[5] defines a POI as a human construct that describes information about locations. Although a POI

[3] Facebook: http://facebook.com
[4] Tripadvisor: http://tripadvisor.com
[5] W3C POI Working Group: http://www.w3.org/2010/POI

is commonly represented as a set of static data (e.g. name, address, geo-coordinates), there are many latent (or hidden) features which describe volatile and temporal aspects of the POI. E.g., the Aug 2011 riots in London constitute an event that temporarily modified its profile from a trendy, fashionable, tourist-friendly location to one characterised by social and political discontent, anarchy and regions of low physical safety. Introducing the ability to model dynamic events enables the provision of both non-changing information and a current view on a POI, filtered based on collective information.

We present *Topica*[6], a mash-up application based on social awareness streams, that enhances the information about a POI by leveraging structured data extracted from the Linked Data (LD) cloud. This is achieved by: (1) removing the need for end users to explicitly contribute information about events and places, hence lowering users' cognitive effort – the mashup extracts this from users' social interaction data; (2) enriching POIs, by identifying relevant concepts in comments related to a POI and matching these to corresponding DBpedia resources. *Topica* demonstrates the use of social awareness streams as a real-time dataset that can be used to reveal dynamic features of a POI. *Topica* provides the following contributions:

1. **Modelling of emerging semantics of POIs based on Social Awareness Streams**
   This stage consists of:
   - *Geo- and time-bounded dataset collection*
   - *Semantic enrichment of social awareness streams*
   - *Triplification of POIs' emerging features*.
2. **Mashup Visualisation of POIs**
   Using the Prism framework[7], *Topica* provides:
   - *Topical exploration of POIs*
   - *Contextual media enrichment*.

2 The *Topica* Application

Modelling of emerging semantics of POIs based on Social Awareness Streams Fig.1 summarises our approach, detailed in the following steps:

1) For a given geographically bounded area, Facebook locations are retrieved;
2) Using the Facebook location properties (e.g. Name, Address, Description) each location is aligned with a Facebook page and a Tripadvisor POI;
3) For each location, comments are extracted from Facebook and Tripadvisor for a time window. The comments are enriched by querying the services *OpenCalais*[8], *Zemanta*[9] and *DBpedia Spotlight*[10], to extract keywords, entities and related pages;
4) This data is used by *Topica* to generate a list of potential DBpedia resources;
5) In order to model the topics in a comment, *Topica* uses the resource list in (4) to query DBpedia. For each resource DBpedia categories (specific and broader) are extracted. E.g., for the resource `Ravioli` these include: `Pasta` and `Italian_`

[6] *Topica* is available at: `http://nebula.dcs.shef.ac.uk/sparks/topica`
[7] Prism Framework: `http://sparksrdf.github.com/prism`
[8] OpenCalais: `http://www.opencalais.com`
[9] Zemanta: `http://www.zemanta.com`
[10] DBpedia Spotlight: `http://dbpedia.org/spotlight`

`Cuisine`, and broader categories: `Mediterranean_Cuisine` and `Italian_Culture`. The set of categories collected from the comments of a page are weighted following a tf-idf (term frequency-inverse document frequency) function;

6) Finally, we introduce the *LinkedPOI* ontology, which represents each POI with a set of static features (the POI's name and geographical coordinates) as well as dynamic features (i.e., topics, tags and messages). The LinkedPOI ontology reuses existing ontologies, including SIOC[11], CURIO[12], and WGS84[13]. This representation allows the use of SPARQL to query for concepts featuring a bounded area[14].

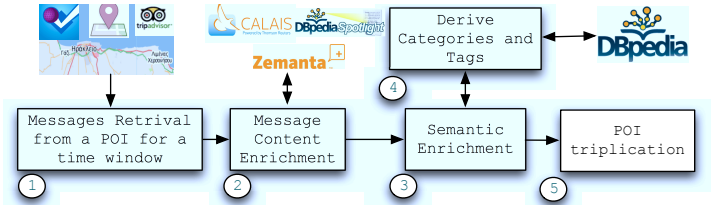

Fig. 1. *Topica* pipeline

Visualisation of POIs The *Topica* visualisation supports the user in retrieving information stored in the LD cloud, by enabling synchronised, semantic filtering of the *Topica* POIs, by taking advantage of the structure defined by the LinkedPOI ontology. LinkedPOI defines a `Patch` as a geographical container of POIs. Each POI has a message container, and each message has tags, entities and category relations. The model of a geographical area is therefore the aggregation of the message containers of the POIs enclosed in the area. Users may perform searches by using filters to trigger seed queries:

1) the *Location lens*, which filters POIs according to their location on a map widget;
2) the *Tag lens*, which enables the selection of POIs according to their associated tags;
3) the *Search lens*, which provides a text filter that operates on the POIs' messages.

A SPARQL query is computed against each set of filter parameters in order to select a subset of the objects returned by the seed query in the visualisation. The user is then able to access the description of a particular POI, by clicking on the corresponding map object. A pop-up window (see Fig. 2) provides a description of the POI: its name and address, tags and topics, the social messages that are associated with the POI and contextual media extracted from the Web.

3 Summary

We have demonstrated a methodology for modelling the collective perception of a POI. We have also illustrated how this model enables topical searches of POIs. *Topica* was

[11] Semantically-Interlinked Online Communities Ontology: `http://sioc-project.org`
[12] Collaborative User Resource Interaction Ontology: `http://purl.org/net/curio/ns`
[13] Geo Positioning vocabulary: `http://www.w3.org/2003/01/geo/wgs84_pos`
[14] SPARQL Endpoint at `http://nebula.dcs.shef.ac.uk/topica/sparql`

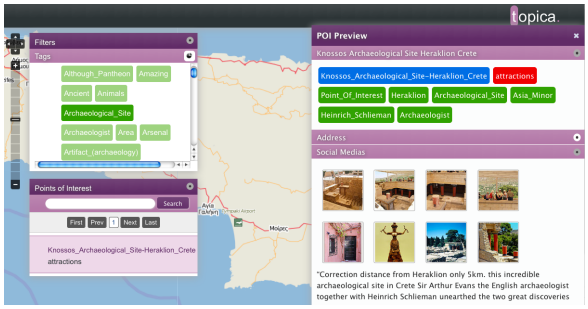

Fig. 2. The Prism *Topica* Visualisation – relevant POIs for the category "Archaeological Sites"

developed to support especially those end users who may have little to no knowledge about where to find information on nearby physical entities, but are able to trigger a search based on their interests. *Topica* caters to the modern user's expectations of ubiquitous technology, by exploiting the collective knowledge of crowds to satisfy overlapping information requirements. The current version uses a fixed dataset for a snapshot in time – Dec 2010 - Jan 2011 – for the island of Crete in Greece. We are collecting data from Bonn, Germany, to allow users at ISWC 2011 to explore this city using *Topica*, according to their topics of interest (e.g. history, night-life, restaurants). An early version of *Topica* placed 2nd in the ESWC 2011 AI Challenge[15].

Acknowledgements A.E. Cano is funded by CONACyT, grant 175203. G. Burel is funded by the European Commission (EC) project Robust (257859). A.-S. Dadzie is funded by the EC project SmartProducts (231204). This work was supported in part by the EC project WeKnowIt (215453).

References

1. F. Abel, Q. Gao, G. Houben, and K. Tao. Semantic enrichment of twitter posts for user profile construction on the social web. In *Proc., Extended Semantic Web Conference 2011*, 2011.
2. S. Asur and B. A. Huberman. Predicting the future with social media. In *Proc., International Conference on Web Intelligence and Intelligent Agent Technology*, pages 492–499, 2010.
3. M. Braun, D. Schmeiß, A. Scherp, and S. Staab. Stevie–collaborative creation and exchange of events and pois on a mobile phone. In *EiMM '10: Proc., 2nd ACM international workshop on Events in multimedia*, pages 35–40, 2010.
4. A. E. Cano, S. Tucker, and F. Ciravegna. Follow me: Capturing entity-based semantics emerging from personal awareness streams. In *Proc, 1st Workshop on Making Sense of Microposts (#MSM2011)*, pages 33–44, 2011.
5. N. Tintarev, A. Flores, and X. Amatriain. Off the beaten track: a mobile field study exploring the long tail of tourist recommendations. In *Proc., 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 209–218, 2010.
6. C. Wagner and M. Strohmaier. The wisdom in tweetonomies: Acquiring latent conceptual structures from social awareness streams. In *Proc. 2010 Workshop on Semantic Search*, 2010.

[15] AI Challenge: `https://sites.google.com/a/fh-hannover.de/aimashup11`