DataTourism: designing an architecture to process tourism data

Fayrouz Soualah-Alila, Mickaël Coustaty, Nicolas Rempulski, Antoine Doucet

L3I laboratory, La Rochelle University

avenue Michel Crépeau, 17000 La Rochelle, France

{fayrouz.soualah-alila, mickael.coustaty, antoine.doucet}@univ-lr.fr

nicolas.rempulski@gmail.com

Abstract

With the rapid diffusion of new technologies in tourism, professionals face new challenges to efficiently use the vast amount of data created by tourists. Nowadays, this type of information comes in huge amount and from multiple and varied sources, such as cellular phones or social networks, touristic location attendance or dematerialized satisfaction surveys. It is an important resource for the tourism industry, but its heterogeneity makes it difficult to aggregate and analyze them. The key issue for tourism actors, professional or governmental decision-makers, is to manage and operate tourism information about their territory effectively. The purpose of this paper is to describe synthetically how tourism information is managed under the *Tourinflux* project. In this paper an architecture named *DataTourism* for tourism data management is described. This architecture solves multiple technological bottlenecks encountered when working with tourism data: heterogeneity, quality, interoperability, reusability and standardization.

Keywords: DataTourism; Tourinflux; TIFSem; TimeML; SentiML.

1 Introduction

In today's rapidly changing world, much data related to tourism is produced. This is primarily the result of increasing possibilities to digitize growing volumes of data, and of the development of open-source and open-data policies. Likewise, more and more data is being generated by sensors, mobile telephones, and connected devices on the one hand, and by the democratization of comparative services dedicated to tourism on the other hand, as Kayak or Yelp for instance. Most of this data could be collected and used by decision-makers to efficiently assign public funds to increase tourist attendance and satisfaction and thus making their territory attractive. But today, they are mostly unused or inefficiently assigned due to a lack of suitable tools.

However, the business sector is already using this data. It is analyzed for marketing strategies, predicting trends and also for producing detailed statistics. Tourism professionals are using multiple sources, and fully use the recent development of the World Wide Web and its social services. The web has changed people's daily life, which is also true for tourism. It has significantly influenced the way information about users is gathered and exchanged in the tourism sector. With the intensive use of social networks and web sites specialized in e-tourism (TripAdvisor, Booking.com, etc.) web users are no longer passive recipients of contents; they absorb information from the web and in return produce their own new content. But when users collect this type of information, from professional sources or from other users, they also

create their own set of information: tourism goods they are looking for, future date of their vacation, etc. Professional tourism services collect this data while providing information or services to users. The same example could be made with cellular carriers, which are tracking movements of their users. This information, wherever it comes from, is then used to improve the service quality by enhancing employee's knowledge about customer's preferences and opinions.

Two main problems occur with tourism data management: their heterogeneity and their volume. As mentioned before, tourism information is continuously enhanced and updated using dedicated websites. These data are contained on web pages that are originally designed to be human-readable, and so, most of information currently available on the web are kept in large collections of textual documents. As the web grows in size and complexity, there is an increasing need for automating time consuming tasks, such as information extraction and interpretation. Some automatic process to annotate and enrich textual information knowledge is thus needed.

The domain of tourism is characterized by significant information heterogeneity and by a high volume of online data. Data related to tourism are produced by different experts (travel agents, tourist offices, etc.) and by visitors, thus creating an heterogeneous data set from a semantic and typology point of view. Moreover, this set is often incomplete and inconsistent. For instance, these data could contain information related to tourism objects (hotels, concert, restaurant, etc.) with raw information, service description for instance, temporal information, about opening hours or days in the week, and opinions, such as users' satisfaction ratings and comments. There are already numerous taxonomies and catalogues which are designed and used internally by tourism actors to allow them to efficiently manage heterogeneous tourism data. Efforts are now made to generate standards to facilitate inter- and intra-tourism data exchange.

The Tourinflux[1] project falls within this context and addresses one central need: to help professional and political actors of the tourism domain to develop the success of their territory. One way to promote territories is to generate reports, also named dashboards, based on enriched data collected from Tourist Information Systems (TIS) and the web. The emergence of dashboards was a consequence of managers' needs to monitor a complex subject with indicators clearly showing how a territory's tourism activities are perceived and evaluated. Experts from tourism industry use and need these dashboards to improve their knowledge about the tourist attractiveness of their territory. But generating dashboards is a problem: as explained above, the heterogeneity in the way information is structured and interpreted leads to conflicts when rich information from different sources needs to be combined. The unstructured nature of data and lack of global schemas means that the available tourism information is human-readable only and not meaningful to machines. Experts from tourism industry are then restricted by the available tools and data structures at their disposal, especially as the task of integrating of heterogeneous data is a time consuming and tedious task to do. Tourism industry needs to access to new tools to increase information suppliers and to easily produce, transmit, access and share knowledge and dashboards.

[1] http://tourinflux.univ-lr.fr/index.php/component/content/?view=featured

The technical architecture created in TourinFlux is aimed at providing the tourism industry with a set of tools (1) allowing them to handle both their internal data, and the information available on the web, and (2) allowing to improve the displayed information available about their territory on the web. In this paper an overall description of how tourism data, composed of information related to tourism objects (TO), temporal information and opinions, are managed under the TourinFlux project, is presented. An architecture named *DataTourism* for the management of tourism data is presented. It allows solving different bottlenecks: heterogeneity of tourism data sources, quality of these data, interoperability, reusability and standardization.

1.1 Designing touristic dashboards

A touristic dashboard is a set of management indicators, built periodically, for a tourist actor or a group of tourist actors, in order to guide their decisions and actions to achieve performance goals. A touristic dashboard is considered as:

- An instrument of control and comparison: It allows tracking the evolution of tourist offers;
- A decision support system to help taking decision: It communicates key information to decision-makers about a touristic activity;
- A communication tool: It provides a permanent communication between the various tourism actors and between different hierarchical levels;
- A monitoring tool: It allows to identify emerging opportunities and risks.

Figure 1 shows a dashboard example from a socio-touristic information report (TourinFlux, 2015).

In France, it exists five main institutional publishers of touristic dashboards: tourist offices, Departmental Tourist Committees (DTC), General Directory of Enterprises of the French ministry of economy (DGE), National Institute for Statistics and Economic Studies (Insee) and the French tourism development agency (Atout-France). Today, each of these publishers has independently developed various techniques to assess a territory (a city, a department, a region, the whole country) and despite all the efforts made so far in developing their own dashboards, these ones have remained insufficient to fulfil the goals described above. More specifically:

- They do not sufficiently represent the tourism activities of a territory. They focus mostly on accommodation and lack other sources of information such as opinion and visitor intents;
- They are limited to the scale of the territory they are developed for. It is impossible to generate dashboards at all hierarchical levels (department, country) or make a comparison between territories. This is mainly due to the heterogeneity of these TIS.

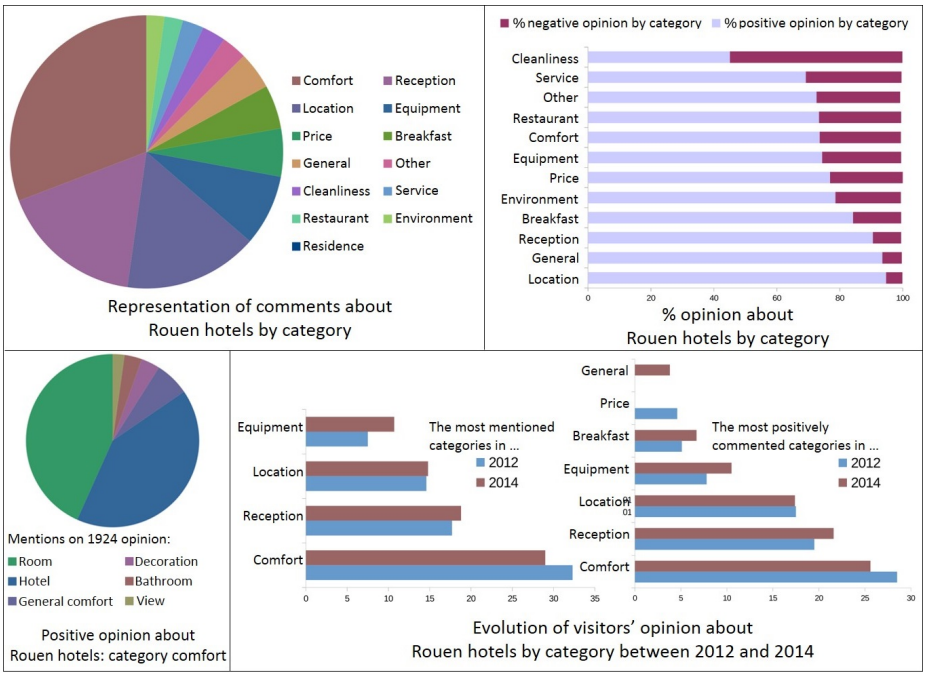

Fig. 1. Example of dashboard

In order to generate rich dashboards, it is necessary to optimally exploit all information available. The data sets used have to be as exhaustive and varied as possible to faithfully reflect the touristic activity of a territory.

In the next section an overview of possible sources and types of tourism data is given, and the current limitations of existing systems are highlighted. Section 3 is devoted to the description of our *DataTourism* architecture for the aggregation of tourism information from different data sources, and the first experimentations led in this way. Finally, section 4 will conclude the paper.

2 e-Tourism issues: sources and types of data

2.1 Sources and types of tourism data

The tourism industry is by nature strongly based on data exchange. In the last decade, more and more data has become available for research and development. This data stems from different sources. The main sources of tourism data which are consider as part of this project are the data available in the different TIS, the data available on the web and open data. These data could be composed of (1) information related to tourism objects (hotels, concert, restaurant, etc.), (2) temporal information and (3) events and opinions.

Our architecture generate dynamic dashboards through four major phases: integration of information related to tourism objects in the system, annotation of temporal information and opinions in web pages, enhancement of the tourism objects from the

annotated information, and finally dynamic generation of dashboards. All the components that are used in each phase are illustrated in Figure 2.

The following sections describe briefly how tourism objects are modelled and how temporal information and opinions in web pages are annotated, so as to complete the description of tourism objects.

2.2 Tourism objects modelling - current limitations

The interoperability of TIS is a major challenge for the development of tourism. Several national, European and international institutional initiatives have proposed different standards to meet the specific needs of tourism professionals, but no international standard have been successfully defined (World Tourism Organization, 2004). French Tourism Actors, in association with the French Ministry of Tourism, created the TourInFrance (TIF) standard in 1999 in order to ease the exchange of tourism data. Major French Tourism Information System such as Raccourci Interactive[2], TourinSoft[3] and Sitra[4], adopted the TourInFrance standard (TIF) at the beginning of the 2000's. It is used today by more than 3,000 tourist offices in France, by DTC and by different tour operators, to facilitate data exchange between these different actors. In 2004, the TourInFrance Technical Group (TIFTG) approved the new version of the standard, TIF V3. In this version, the standard has evolved towards XML technologies to facilitate the publication of information on the web and the exchange of information between systems. It comes together with several thesaurus. Since 2005, this standard stopped evolving. As a result, tourism professionals have adapted the standard to their own needs (new tags added, varying syntax, etc.) and proposed their own evolution in an unorganized way. With the creation of web technologies and the democratization of open-data, this standard became outdated and TIS lost their inter-compatibilities and cannot directly share their data using international standards. Finally, the lack of international standards, in accordance with the exploitation of tourism information, are trapped in their own territory, and thus it is a complicated task to aggregate these information (Bittner et al., 2005).

Based on this inventory, the next section presents a new system for tourism data management. The challenge confronted in this paper imposes two major restrictions. First, the designed system must be able to model and structure knowledge from the domain of tourism and those from the lingual domain, while being inter-connectable with semantic existing systems. On the other hand, this system must address the problem of big data as a huge volume of opinion data are produced every day on the web, and ontologies may be overpassed by the quantity of data to perform and to share. This model must then be able to deal with the need of velocity. We therefore propose to use a combination between dedicated annotation languages to be able to treat quickly large corpus, while ontologies give a global framework to structure data and to ensure their inter-compatibilities. Finally, this system can be easily linked to web semantic technologies in order to ease the production of dashboard and the exchange of data (like dashboards for tourism actors).

---

[2] http://www.raccourci.fr/

[3] http://www.tourinsoft.com/

[4] http://www.sitra-tourisme.com/

3 Overview of the proposed approach

To overcome these limitations, a proposition would be to evolve the TIF standard to share the knowledge it contains and to ensure data interoperability, by applying the concept of ontology to represent the standard terminology. This new standard is based on the French initiative, but is not limited to French data. All its components remain generic and can be easily adapted to the international market.

Ontologies have been chosen as they are defined as "the specification of a conceptualization", in other words, as "a specific artifact designed with the purpose of expressing the intended meaning of a shared vocabulary" (Hirst, 2004).

Having a common semantic base alleviates the interoperability bottleneck (Fodor and Werthner, 2005) that comes along with the integration of heterogeneous data sources by converting existing heterogeneous unstructured tourism data into structured ontological data. In the tourism area, some research work has already tackled the design of ontologies. Several available tourism ontologies show the current status of the efforts: the OTA (Open Travel Alliance) (OTA, 2000), the Harmonise ontology (Dell'Erba et al., 2002), the Hi-Touch ontology (Legrand, 2004), the QALL-ME ontology (Ou et al., 2008), the Tourpedia catalogue (Cresci et al., 2014), etc. These models focus on different areas of the tourism domain, but none of them deals with all the areas together, hence failing to provide an overview of the data required for a full dashboard. To the best of our knowledge, no unique ontology exist to overcome this problem.

As shown in Figure 2, the global organization of the proposed standard relies on a modular framework, and is actually composed of three main components:

1. An evolution of TIF into TIFsem in order to store tourism data in a format compatible with the semantic web technologies so as to ease the sharing and the search of data;
2. An evolution of the TimeML standard in order to adapt it to the specificities of temporal data;
3. An evolution of the SentilML standard in order to be able to deal with opinion data.

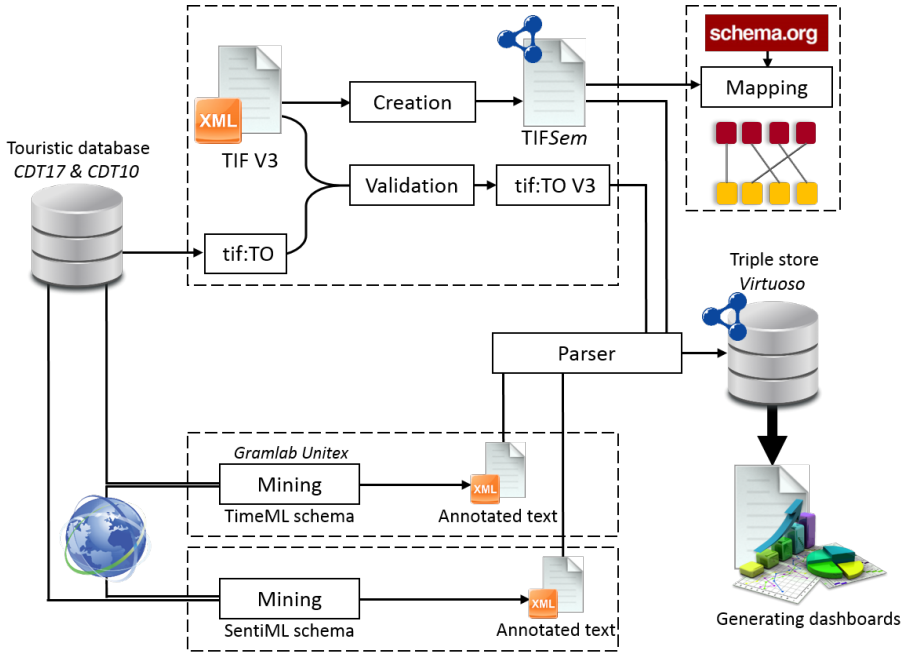

Fig. 2. DataTourism general architecture

3.1 Tourism data standard

As mentioned previously, no international standard actually exist for tourism data exchange and reasoning. This standard must be able to offer two kinds of services (knowledge extraction and reasoning and knowledge sharing), and to deal with heterogeneous information (textual information, GPS position, meteor or temporal data, etc.). An heterogeneous solution is proposed combining ontology for the structuration and the reasoning part, and a full-compatibility with Schema.org[5] formalism, internationally recognized for its ability to share and to spread knowledge.

For the first part of our model, an ontology called TIFSem (Semantic TourInFrance) is proposed to globally describe tourism objects mixing heterogeneous content (Soualah-Alila et al., 2015). We chose to re-design the TIF standard in TIFSem for reasoning purposes. An ontology for tourism domain implements mechanisms of deductive reasoning, automatic classification, information retrieval, and ensure interoperability between TIS. Concepts included in the defined ontology will allow to describe information sources on tourism. This model allows enriching the tourism information to be used: (1) from the user side, to match tailored package holidays to client preferences for instance, and (2) from tourism experts' point of view, to analyze and better manage online data about their territory. As tourism dashboards require to analyze vast amount of data, reasoning can lead to better indicator. For instance, it can help profiling tourist based on their behavior in an area with simple set of rules.

[5] https://schema.org

In order to elaborate the TIFSem ontology, we exchanged with different tourism actors to understand their domain, their specific data and create concepts related to this specialized domain. We collaborated with sources from the Departmental Tourism Committee of the Charente Maritime[6] (CDT17) and the Departmental Tourism Committee of the Aube[7] (CDT10). We are also in the process of extending the TIFSem ontology by collecting content from more tourism service providers abroad from France.

Our second goal is to ensure that TIFSem is compatible with current data crawling and data publication technologies. As TIF standard is unable to easily share and interoperate with global standards from the Web, we propose to enrich the TIFSem ontology with the Schema.org model. This schema, initiated few years ago by Bing, Google and Yahoo, aims to standardize structured data formats of the semantic web and is a *de facto* norm to easily share semantic content in the web. Launched in 2011, Schema.org aims to create and to support a common set of schemas for structured data mark-up Web pages in a way recognized by major search providers, and that can also be used for structured data interoperability (RDFa, JSON-LD, etc.). When these tags are used in a website, search engines can better interpret the meaning of its embedded resources (text, image, video) (Toma et al., 2014). The proposed model tends to match terms of TIFSem with terms of Schema.org by using semantic relations. Moreover, by working with the Schema.org community, we intend to extend the schema, either formally by adding new terms or informally by defining how Schema.org can be combined with some additional vocabulary terms.

The global model obtained by mixing these two tools is presented in Figure 2. In order to feed this model, the next sections present how we interfaced it with automatic processing techniques for the extraction of basic information, related to the tourism domain (time, opinions). This allow us to start from documents, extract some keywords and annotations, insert them into our model to finally infer new knowledge and provide higher level of semantic. In the following sections we describe how in our architecture we integrate information about time and opinions to complete the TIFSem model

3.2 Tourism temporal data

The first extension of the TIFSem model relies on the use of temporal information in tourism corpora. Temporal data are pieces of information frequently encountered in tourism web pages. Most tourism objects (events, hotels, restaurants, etc.) on the web are associated to periods and events and are characterized by different timestamps like date, duration, opening hours, opening conditions, frequency, etc. Textual tourism data on the web is a rich body of phenomena for linguistic analysis. The automatic recognition of temporal and event expressions in natural language text has recently become an active area of research in computational linguistics and semantics.

Temporal annotation is an essential part of many text understanding efforts. Recent efforts such as TIDES (Translingual Information Detection, Extraction, and Summarization) (Ferro et al., 2001), STAG (Sheffield Temporal Annotation

[6] http://www.charente-maritime.org/
[7] http://www.aube-champagne.com/

Guidelines) (Setzer, 2001) (Setzer and Gaizauskas, 2001) and TimeML (Pustejovsky et al., 2005) all aim to provide a markup language for temporal annotation. TIDES defines a set of guidelines for annotating time expressions with a representation of the times they refer to. STAG provides guidelines for annotating events and temporal information in newswire texts. TimeML is an extension of TIDES and STAG. In our proposed model, events will be annotated according to the TimeML language, a robust specification language for the challenging task of annotation of temporal information over natural language text. Under the TourinFlux project, a state of the art has been proposed by (Drat, 2014) in order to justify the use of this language.

TimeML has been developed in the context of AQUAINT workshops and projects. The 2002 Time and Event Recognition for Question Answering Systems (TERQAS) workshop set out to enhance natural language question answering systems to answer temporally-based questions about the events and entities over free text on the web. This is when the first version of TimeML was defined and the TimeBank corpus created as an illustration. In 2003, TimeML was further developed in the context of the TimeML Annotation Graphical Organizer (TANGO) workshop. In 2009 TimeML has been developed into an ISO standard (ISO WD 24617-1:2007).

TimeML includes four major data structures: EVENT, TIMEX3, SIGNAL, and LINK. In TimeML, *events are situations that occur or happen, or predicates that describe states or circumstances in which something obtains or holds the truth* (Pustejovsky et al., 2003). Events in TimeML are annotated with the tag EVENT. TIMEX3 is used to tag explicit temporal expressions, such as time, dates, and durations. SIGNAL is used to annotate sections of text, typically function words that indicate how temporal objects are related to each other (when, during, before, etc.). Finally, LINK encode various relations that exist between the temporal elements of a content. Each of these tags are associated to attributes to integrate temporal expressions. As the description of TimeML is not the aim of this paper, a complete description of the language is given in its manual (Sauri et al., 2009).

Within the Tourinflux project, in order to facilitate the extraction of temporal data, a corpus of Web pages linked to tourism is created in the purpose of being analyzed. This corpus consists of:

- A free text corpus containing festivals and events description, provided by the Local Action Group of *Othe Armance*[8]. This corpus is available under LGPL/LR license (Lesser General Public License for Linguistic Resources);
- A corpus provided by the CDT10. This corpus contains descriptions of Places of Interest (POI): hotels, restaurants, etc., In particular, it contains information about opening and closing dates, opening and closing times, etc.;
- Open data, including data concerning national museums.

We perform the annotation of our corpus annotation with temporal expressions with a set of finite state transducers, developed with the Gramlab Unitex[9] corpus processor. Unitex is a corpus processing system for analyzing natural language texts using resources such as dictionaries and grammars. Gramlab is an integrated development

[8] http://www.tourisme-othe-armance.com/
[9] http://www-igm.univ-mlv.fr/~unitex/

environment, based on the Unitex software components, designed for industrial project management purpose. Before applying the transducers, Unitex performs some pre-processing that consists in cleaning the text, by (1) normalizing apostrophes, quotes etc., (2) segmenting the text into sentences and tokenizing it and (3) applying a number of built-in lexical resources, such as dictionaries to identify, for instance, compound word forms, proper names, etc. (Paumier, 2008). Once the text is cleaned, temporal expressions are tagged according to their TimeML type. The tagger performs the identification of events. Then Unitex detects and annotates temporal expressions and calculates the attribute value for each of the tags as specified by the TimeML guidelines. The tagger also detects certain relation markers, such as temporal prepositions like before, after, etc. The last spot of the tagger is to determine the links between the different annotations. The resulting output of Gramlab is the original corpus annotated with EVENT, TIMEX3, SIGNAL and LINK tags, whose values can later be integrated within the TIFSem model. The results of our work on opinion annotation with SentiML are presented in (Drat, 2014).

3.3 Tourism opinion data

An important part of our information-gathering behavior has always been to find out what people think about their touristic experience. Opinions help to analyze a situation from different aspects and take an appropriate decision. The opinion of one individual may influence another individual's opinion and hence the concept of public opinion is generated. Public opinion is very important in the tourism domain.

The amount of opinionated data on tourism websites has exponentially increased especially after the rapid growth of online social networks. With the availability and popularity of rich opinion resources, we need to have reliable mechanisms to identify all aspects of opinion in a text and extract useful related information. Thus, we introduce the concept of opinion mining.

Opinion Mining is the process of automatic extracting opinions from textual segments (Liu, 2012). In the literature, it has commonly been referred to as sentiment analysis or sentiment classification and sometimes as subjectivity analysis (Cambria et al., 2013). There are many related sub-tasks of opinion mining, such as the semantic annotation of opinions. Semantic annotations are essential both to prepare data for machine learning and to evaluate opinion mining approaches. Some annotation schemas have been proposed by the research community such as SentiML (Di Bari et al., 2013), OpinionMining-ML (Robaldo and Caro, 2013) and EmotionML (Schröder et al., 2011). A comparative study between the existing annotation schemas is presented by (Malik et al., 2014) as part of the project TourinFlux.

In our case, we used SentiML for annotating opinion data. In SentiML we talk about sentiments rather than opinions. The goal of SentiML is to identify and classify sentiment groups (positive and negative) at the sentence level. In order to do this, the schema focuses on three categories: target (expression the sentiment refers to), modifier (expression conveying sentiment) and appraisal. A target is any entity (object, person or concept) that is implicitly or explicitly regarded as positive or negative by the author of the text. A modifier is what modifies the target. It can be an adjective, a verb, an adverb or a noun. However, SentiML also adds in its vocabulary the much needed appraisal tag. An appraisal group represents an opinion on a specific

target. For this reason, it is defined as the link between the target and the modifier (e.g., link between a noun and an adjective, or between a verb and an adverb, etc.). Besides this, SentiML is based on the Appraisal Framework (AF) which is a strong linguistically-grounded theory. AF helps to define appraisal types (affect, judgments and appreciation) within the modifier tag. The results of our work on temporal annotation with TimeML are presented in the report of (Malik et al., 2014).

4 Conclusion

In this paper, we presented some early stage work in the Tourinflux project on identifying a new architecture named *DataTourism* for tourism data management. We described in a general way how tourism data are managed to help experts from tourism industry to generate dashboards to improve their knowledge about the tourist attractiveness of their territory. This architecture is being built. The three main components of the architecture (the evolution of TIF into TIFsem, the annotation of temporal data with TimeML and the annotation of opinion data with SentiML) has been validated. The overall framework is being tested and evaluated with partners. One of the limitations of the project is that it is actually restricted to a national level. We are also working on extending *DataTourism* for managing tourism data at an international level. We are also in the process of extending the TIFSem ontology by collecting contents about more touristic service providers. We are also working on us cases for generating dashboards and providing semantic and contextual answers to queries.

References

Bittner, T., Donnelly, M. & Winter, S. (2005). Ontology and Semantic Interoperability. Zlatanova, S & Prosperi, D. (Eds), *Large-Scale 3D Data Integration: Challenges and Opportunities*, 139-160.

Cambria, E., Schuller, B., Xia, Y. & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28(2), 15-21.

Cresci, S., D'Errico, A., Gazze, D., Duca, A. L., Marchetti, A. & Tesconi, M. (2014). Towards a dbpedia of Tourism: the Case of tourpedia. *International Semantic Web Conference*, 129-132.

Dell'Erba, M., Fodor, O., Ricci, F. & Werthner, H. (2002). Harmonise: A Solution for Data Interoperability. *Towards the Knowledge Society: eCommerce, eBusiness, and eGovernment, the Second IFIP Conference on E-Commerce, E-Business, E-Government*, 433-445.

Di Bari, M., Sharo□, S., Thomas, M. (2013). Sentiml: Functional Annotation for Multilingual Sentiment Analysis. *1st International Workshop on Collaborative Annotations in Shared Environment: Metadata, Vocabularies and Techniques in the Digital Humanities*, DH-CASE'13, ACM, New York, USA, 15:1-15:7.

Drat, L (2014). Projet TourInFlux. Annotation des Expressions Temporelles.

Ferro, L., Mani, I., Sundheim, B., & Wilson, G. (2001). TIDES Temporal Annotation Guidelines, Version 1.0.2. *The MITRE Corporation, McLean, Virginia. Report MTR 01W0000041*.

Fodor, O. & Werthner, H. (2005). Harmonise - a Step Towards an Interoperable e-Tourism Marketplace. *International Journal of Electronic Commerce*.

Hirst, G. (2004). Ontology and the lexicon. Staab, S. & Studer, S. (Eds), *Handbook on Ontologies*: Springer-Verlag, 209-229.

Legrand, B. (2004). Semantic Web Methodologies and Tools for Intra-European Sustainable Tourism. White paper, Paris, Mondeca.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers.

Malik, M., Missen, S., Attik, M., Coustaty, M., Doucet, A. & Faucher, C. (2014). SentiML ++: An Extension of the SentiML Sentiment Annotation Scheme. *The 12th Extended Semantic Web Conference (ESWC2015)*.

Ou, S., Pekar, V., Orasan, C., Spurk, C. & Negri, M. (2008). Development and Alignment of a Domain-Specic Ontology for Question Answering. Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S. & Tapias D. (Eds), *the Sixth International Language Resources and Evaluation Conference*, 2221-2228.

OTA (2000). Opentravel Alliance. Opentravel Alliance message specications. Specications Document 1.

Paumier, S. (2008). UNITEX 2.0: User Manual.

Pustejovsky, J., Castano, J., Ingria, R., Sauri, R., Gaizauskas, R., Setze,r A. & Katz, G. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. *IWCS-5 Fifth International Workshop on Computational Semantics.*

Pustejovsky, J., Ingria, B., Sauri, R., Castano, J., Littman, J., Gaizauskas, R., Setzer, A., Katz, G. & Mani I. (2005). The Specification Language TimeML. *The Language of Time: a Reader,* 545-557.

Robaldo, L. & Caro, L. D. (2013). OpinionMining-ML. *Computer Standards & Interfaces*, 35 (5), 454-469.

Sauri, R., Goldberg, L., Verhagen, M. & Pustejovsky, J. (2009). Annotating Events in English, TimeML Annotation Guidelines, Version TempEval-2010.

Schröder, M., Baggia, P., Burkhardt, F., Pelachaud, C., Peter, C. & Zovato, E. (2011). Emotionml - an Upcoming Standard for Representing Emotions and Related States. *Affective Computing and Intelligent Interaction*, Springer.

Setzer, A. (2001). Temporal Information in Newswire Articles: an Annotation Scheme and Corpus Study. Ph.D. thesis, University of Sheffield, Sheffield, UK.

Setzer, A. & Gaizauskas, R. (2001). A Pilot Study on Annotating Temporal Relations in Text. *ACL 2001, Workshop on Temporal and Spatial Information Processing.*

Soualah-Alila, F., Faucher, C., Bertrand, F., Coustaty, M. & Doucet A. (2015). Applying Semantic Web Technologies for Improving the Visibility of Tourism Data. *CIKM'15 Workshop on Exploiting Semantic Annotations in Information Retrieval ESAIR'15*.

Toma, I., Stanciu, C., Fensel, A., Stavrakantonakis, I. & Fensel, D. (20014). Improving the Online Visibility of Touristic Service Providers by Using Semantic Annotations. *The Semantic Web: ESWC 2014 Satellite Events*, Anissaras, Crete, Greece, 259-262.

TourinFlux (2015). Ville de Rouen: Rapport d'informations socio-touristiques.

World Tourism Organization (2004). Information and Documentation Resource Centres for Tourism: Guidelines for Establishment and Maintenance. ISBN: 9284407176, 9789284407170, 132.