ConTur: an intelligent content management system for the tourism sector

Germán Herrero Cárcel[a], Aida Campo[b], Maria Gil[b], Ana Garcia[c], David Martin, Ion Zugasti[d], Sonia Bilbao[e], Amalia Perez[f], Hristo Koshutanski[g], Antonio Maña[g] and Iñaki Perez De Albeniz[h]

[a] Atos Research & Innovation, Atos Spain, Spain
[german.herrero]@atos.net

[b] Andago Labs,Andago, Spain
[aida.campo, maria.gil]@andago.com

[c] Dome Consulting, Spain
[agarcia]@dome-consulting.com

[d] Cooperative Research Center in Tourism, CICtourGUNE, Spain
[david.martin, ion.zugasti]@tourgune.org

[e] Tecnalia Corporacion Tecnológica, Spain
[sonia.bilbao]@tecnalia.com

[f] Universidad de Burgos, Spain
[apromero]@ubu.es

[g] Computer Science Department, Universidad de Malaga, Spain
[hristo,amg]@lcc.uma.es

[h] Barcelo Viajes, Spain
[i.perez]@barceloviajes.com

Abstract

Nowadays, tourists have access to a lot of different web sites in order to find information about destinations. This has a direct impact on tourism destination management organizations and tourism providers. These stakeholders have a lot of problems in order to find reliable and up-to-date information about their destinations. This paper presents an innovative solution that helps to promote tourism destinations through an intelligent, reliable and comprehensive content management platform that integrates different online heterogeneous and distributed information sources. This platform is described and validated in two different scenarios: "Portal Destino 3.0", a tourism portal managed by a public organization and "Mi Guia Multidestino", a personalized online web travel guide.

Keywords: intelligent content management system, semantic annotation, intelligent travel guides, semantic content integration.

1 Introduction

Although often underestimated as a marginal sector, the tourism sector plays a strategic role with a considerable market share for the European continent. Europe is still the leading tourism market worldwide with more than half of the global market

share (488 million arrivals) in 2008[1]. The importance of Tourism as a strategic market sector is also underpinned by its share of the European GDP (5% of the total European GDP is produced directly and 10% of it indirectly by Tourism) as well as the number of jobs depending on it (7-8M jobs directly and 20M jobs indirectly). However, the tourism sector has its strengths and weaknesses. Some of the most relevant weaknesses in the European and Spanish tourism are:

- Emerging destinations growing rivalry (problems of competitiveness)
- Weak sector integration: the value-chain of the tourism sector is weakly integrated due the emergence of Internet and the fragmentation of tourism service providers.
- Nowadays, one of the trends that has had most effect on the sector in the recent years is the ever-growing number of website pages providing tourism information and the use of social-media that is starting to pervade every single internet services and defines the online user behaviour and consumer decisions

To solve the above-mentioned issues is more necessary than ever to develop a system that helps to promote tourism destinations through and efficient management of content partners and integrate the actors of the tourism value-chain related to the destination. ConTur faces these challenges and its main objective is to provide an intelligent, reliable and comprehensive management of different online heterogeneous resources (personal websites, social media, documents and other resources). ConTur platform has a global reach within the tourism sector. The benefits provided and the problems it intends to address concerns all actors which are part of the value chain (public tourism authorities, service providers, online agencies, tourists).

2 Background

Currently, the World Wide Web (WWW) is based mainly on documents written in HTML, a mark-up language used to create hypertext on the Internet. The Semantic Web aims to overcome the limitations of HTML by introducing explicit descriptions of meaning, the internal structure and the overall structure of content and services available on the WWW (Berners-Lee, 2001). For this, the Semantic Web technologies mainly used are RDF, SPARQL, and OWL, as well as XML, mechanisms to help turn the Web into a global infrastructure where it is possible to share and reuse data and documents between different types of users (W3C).

In the particular case of the tourism sector, where different kinds of information, provided by different sources are generated, ontologies can be applied to improve interoperability between different sources provided by the actors that take part in the tourism sector. Ontologies act as another tool in the content management systems or knowledge management. They transform the information, and increase it with structures that give content and meaning to information.

Moreover, at present, exchanging information between companies is based on the different standards. These standards have been used as agreements between different actors in a given sector to ensure that components made by different organizations can

---

[1] European Travel Commission – Fact Sheet, September 2009:

http://www.etc-corporate.org/resources/uploads/ETC_Factsheet_20090831.pdf

be mounted on a set and work in coordination. After several years of standardization, leading companies are beginning to adopt Service-Oriented Architecture (SOA) to integrate disparate data, providing the availability of data using Web services. Web services provide standardized and open protocols to provide a means unified access to information from a diverse set of platforms (different operating systems, programming languages, and applications). These Web services can be reused to provide new services and applications within the same organization or between different, generating a high flexibility for the business.

Finally, a key aspect of the current web is the emergence of the concept of *mashup*. Mashups technologies can be used through the philosophy of Software as a Service (SaaS) (Piñero, 2007). A mashup is a web application that uses resources from more than one source to create a complete service. The data used by the *mashup* is normally obtained from third parties, accessing public interfaces or APIs they provide, through interoperable items, which are integrated through syndication tools.

Thus, it can be concluded that the application of technologies based on the Semantic Web can improve access to tourist information integrated from multiple sources of information scattered (Murua, 2005).

3 Objectives

The main objective of ConTur is to develop a Content Management System for tourism destinations to enable tourism stakeholders (e.g. SMEs) to be present on the Internet directly to travellers or receptive businesses. ConTur platform integrates content from different sources: tourism companies at destination, unstructured information on the Internet (e.g. social media, blogs), DMOs and local institutions.

With ConTur, any stakeholder can filter, catalogue, aggregate, merge and integrate tourism related contents coming from different, heterogeneous and distributed web sites (e.g. private web sites, social networks, institutional catalogues, etc.) in an intelligent, reliable and robust way.

The scope and purpose of the platform is multiple and varies depending on the stakeholder of the tourism value chain. In the case of the institutional agencies of tourism, its aim is to encourage the access to their content (e.g. brochures, videos, cultural agendas, etc.) and the promotion of the destination itself through different indirect channels (e.g. online agencies, social network of travellers, etc.). Similarly, they may rely on the platform to enrich their portals with the inclusion of non-institutional content available on the Web. Destination service providers will increase direct sales channels due to the contents being captured in the Web, and will facilitate the dissemination and promotion of its services through indirect channels. In the case of online agencies, they will be able to offer value added services, based on the creation of personalized contents. The offer will become more attractive thanks to the inclusion of content from the destination. Finally, tourists, consumers and producers of the tourism sector, will be able to choose their destination and prepare a more exhaustive travel plan thanks to the information that is acquired from different sources.

4 ConTur Architecture

ConTur has to acquire and unify content from different sources that is usually structured using different data formats. This way, the proposed architecture has to face the following technological challenges:

- Data structures that facilitates the description of tourism contents through a tourism ontology.
- Web interfaces to capture and annotate tourism content (automatically and semi-automatically).
- Mapping mechanisms that allow the transformation of content's metadata already described by their suppliers to the ConTur ontology.
- Ontology extension mechanisms to facilitate the identification of new tourism concepts that were not initially included in the ontology.
- Reasoning engines to classify and infer new data based on the acquired contents.
- Content aggregation engines that fuse content selectively according to the requirements of the recipient.
- Web interfaces in order to facilitate the visualization of content by different users.
- Security and confidence models that allow the definition of access policies and the allocation of degrees of reliability to the content.
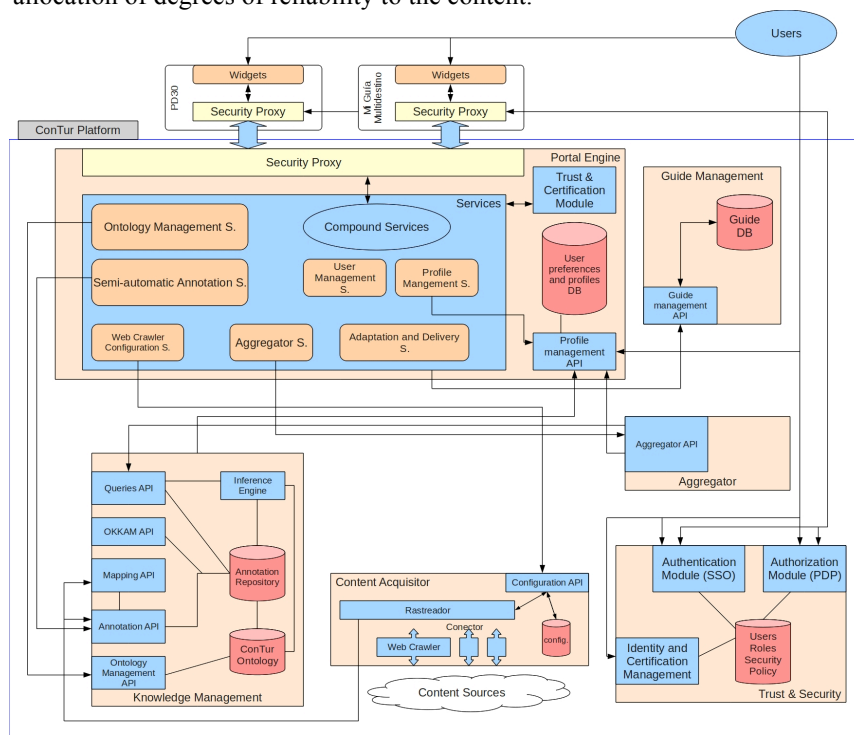

Fig. 1. ConTur Architecture

As shown in Fig.1, ConTur is divided into six different modules. Each of these modules is described in the following sections.

4.1 Portal Engine

The portal engine is the entry point to the platform. Here is where services basic and compound services are found, so that users from different use cases have the opportunity to access the functionality offered by the platform.

ConTur platform contains a Content Management System (Generic Access Portal) from which widgets that can be embedded in other web pages can be downloaded. These widgets are web user interface components that allow the communication with the platform's services. Each widget communicates with a compound service, which can communicate with other simple services (or *mashups* of simple services). Communication between widgets and services is carried out through SOAP (Simple Object Access Protocol) web services and through security proxies.

In addition, in this Content Management System, a management interface for administrators is provided, and includes the following functions: registration of new users, configuration of the crawler (including the web pages relevant to the tourism domain and other parameters), extension of the ontology, etc. Beta version of Portal Contur (the Generic Access Portal of ConTur's project) can be accessed at this URL: http://portalcontur-demo.andago.com/

4.2 Knowledge management

This module manages all the information that is gathered from tourism web sites. This way, the information that is annotated by the users of the platform is sent to the Knowledge Manager. Here, this data is saved and merged into the triple store, avoiding data inconsistencies and redundancy. This information can be queried in order to be used by the rest of the modules of the platform.

Data is structured in RDF format according to the defined semantic model. Two different stores have been configured. The first one contains the structure of the ontology (TBox) and the second one contains all the annotations in the form of triples (ABox). The module has been implemented using the Jena semantic framework .

The Knowledge Manager is divided into several modules that expose high level APIs to the rest of the platform.

- Ontology Management API: it is used to extend the proposed ontology. Like that, new classes and properties can be created and modified. This module has an extra store where all the created/modified classes and properties are stored in order to be validated by the administrator.
- Annotation API: this module receives the created annotations and inserts them into the repository of annotations. It uses the Okkam entity name system (http://www.okkam.org) in order to assign or reuse unique identifiers to the triple instances.
- Mapping API: this is used to create mappings between web contents that are already structured using ontologies and the ontology used by the platform.
- Query API: it is used to retrieve data from the Knowledge Manager using the SPARQL query language.

- Inference engine: this module is used to make inferences based on the asserted triples and the defined rules. The inference process generates new triples that are also stored in the repository of annotations.

The created ontology has been designed according to the state of the art in the field of tourism semantic modelling (Prantner, 2007) (Werthner, 2009). This ontology is defined using the OWL ontology language and contains 153 classes and 116 properties. It uses other ontologies like the w3c time (http://www.w3.org/2006/time) and geo (http://www.w3.org/2003/01/geo/wgs84_pos) ontologies in order to model the time and the location entities.

4.3 Content Acquisitor

This component is responsible of collecting touristic contents stemming from different and heterogeneous sources scattered around the web, such us private web pages, social nets, institutional catalogues, etc. and annotate them automatically based on the ontology of ConTur. Content *acquisitor* also includes semi-automatic annotation system to accomplish editorial revisions over annotations performed automatically, which combines the active learning and self-training in order to reduce the efficiency of the annotation process.

The Content *acquisitor* is composed by the following modules; some interact with APIs that reside inside ConTur platform:

- Crawler configuration API: This component allow to ConTur administrator to set parameters to configure the crawler, which are stored in DB configuration. It can be set the seed URLs from which the crawler begins or the refreshing interval to re-crawl the pages among other properties.
- Crawler: This process crawl the web to recollect sources of unstructured format (mainly HTML) with tourism value. It starts from the seed URL's and crawl connected pages so long as they continue to be considered relevant to tourist domain topic. The relevance of a page is determined after downloading its content. To calculate the relevance to a predefined set of topics web documents are analyzed by annotation API. Relevant pages are sent to content indexing and their contained URLs are added to the crawl frontier, pages that fall below are discarded. ConTur Crawler has been implemented by building a focused (vertical) crawling mechanism on top of Nutch open source crawler, which uses MapReduce distributed model to make the crawler scalable.
- Text processing: Text is processed to detect semantically related concepts already present in ConTur domain ontology. It uses annotation API to do the Natural text processing.
- Annotation repository: Extracted touristic concepts are annotated against ConTur domain ontology and annotations are stored on ConTur repository by calling annotation API.
- Semi-automatic annotation API: This component provides results from automatic annotation process. A user can inspect these results and refine or correct them if required.

The semi-automatic annotation API is based on Freeling API [Freeling, 2010], an open source suite of language analyzers. The annotation API receives the corpus from

the crawler and executes a natural language process through a text tokenization and entity detection. When the entities are identified, the API produce a query against the repository to check if the entity already exists, in order to have the URI of the semantic entity from the ConTur ontology. The new annotation is added to the repository in the form of a triple as follow:

{<URLsource>, <ConTur:isAbout>, <ConTur:SemanticEntityfound>}.

If the entity identified in the corpus is not in the repository, the Annotation API uses the Okkam entity name system in order to assign a unique identifier to the triple instance before add the annotation to the repository. Based on Freeling capabilities, the current version of the Annotation API supports languages as Spanish, English, Italian or Portuguese. Ontologies are a good modelling approach in order to provide a common vocabulary and harmonize information coming from heterogeneous sources. Furthermore, reasoning techniques can be applied in order to infer new knowledge from explicit information that can be valuable for the rest of the platform.

4.4 Aggregator

As the portfolio of products and services of a destination is so numerous, it becomes increasingly important to offer tools that help travellers find those that match their demand and expectations.

The aggregator is a search engine that returns the set of tourism products and services that match the user preferences as expressed in a query. It has been designed as a configuration tool of products and services. By configuration we understand the definition given by Mittal & Frayman [Mittal et al, 1989].

> *"Given: (A) a fixed, pre-defined set of components, where a component is described by a set of properties, ports for connecting it to other components, constraints at each port that describe the components that can be connected at that port, and other structural constraints (B) some description of the desired configuration; and (C) possibly some criteria for making optimal selections.*
>
> *Build: One or more configurations that satisfy all the requirements, where a configuration is a set of components and a description of the connections between the components in the set, or, detect inconsistencies in the requirements."*

In the case of the tourism domain the components can be hotels, holiday packages, flights, restaurants, car rentals, etc. Each component is described by a set of properties such as the name, rating, popularity, price, brand, amenities and so on.

Components are associated or "connected" to a destination and to other resources. These resources can be multimedia resources (audio, photo or video) or websites (blog, social network or official website).

A semantic model has been defined in order to translate into a machine readable description the query performed by the user. The tourist preferences are expressed as constraints on the components properties. These constraints are used to filter the amount of results. In addition, the semantic model allows defining ranking criteria to sort the results returned by the aggregator. Finally, the model allows specifying the data to be retrieved for each type of component that is part of the query's results, e.g.

name and address for hotels, description for activities and price for car rentals. [Bilbao, 2010].

4.5 Guide Management

This module provides the CRUD functionality (Create, Read, Update, Delete) to manage travel guides. Each travel guide is composed by a list of destinations and a period of time in which to visit each destination. By means of the aggregator, the user can search for tourism products and services associated to each destination. If the product returned is of interest for the user, he can add it to his travel guide. Users can decide the level of access security for each of their guides: private, protected or public.

4.6 Trust & Security

The ConTur platform is designed to offer touristic contents to end users via a federation of Web portals. Given that, it was developed a security architecture providing a *circle of trust* between Web portals and the ConTur platform, where both Web portals and ConTur platform have to be certified by a ConTur certification authority to be part of the circle of trust.

The security solution builds upon single sign-on (SSO) for federations and provides single authorization for federated content provisioning. It was adopted SAML (SAML, 2005) specification for building SSO-based federated identity management capable of expressing various authentication contexts such as X.509 (X.509, 2005) certificates – the case of FNMT[2] and DNIe[3] certificates, or credentials of type userID/password.

A *single authorization* protocol allows for multiple unified user accesses to Web portals within the ConTur federation. Users get authorized for service access once (within an SSO session) and for each Web portal of the federation obtain interface access without being required to authorize for the already authorized services.

A ConTur Security Provider (CSP) actor implements "Trust & Security" functionality of Fig.1 as logically independent entity from the platform. The separation facilitates management of security aspects from those of the business logic of the platform. The CSP serves as an anchor of trust for the federation. Web portals trust the CSP for proper user authentication and authorization when users interact with Web portals' interfaces, while the ConTur platform trusts the CSP for certifying ConTur Web portals and providing authorization statements about end-users when Web portals access platform's services on behalf of end users.

Fig. 2 shows a high-level view of security interactions between the main actors. A security proxy was developed to facilitate secure and trusted communications between Web portals and CSP, and between Web portals and the ConTur platform.

---

[2] http://www.fnmt.es

[3] http://www.dnielectronico.es

The proxy encapsulates all authentication and authorization aspects providing high-level security abstraction.
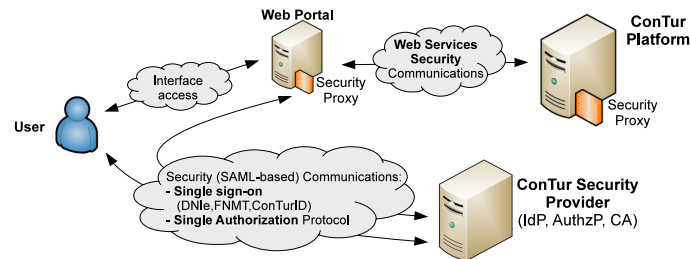

Fig. 2. ConTur Security Communications

5 Scenarios

5.1 Portal Destino 3.0

The main goal of the "Portal Destino 3.0" is to allow the management, access and display of contents of the destination. This site presents the user, through the use of dialog boxes, the contents automatically processed and added by the content manager, whose sources include institutional and non institutional content. This tourism portal is based on the integration of existing information distributed by other portals, institutional or otherwise, on tourism and related topics. For its construction, data sources provided by ConTur platform are used. In this case, ConTur provides a mechanism that allows the integration of automated information on tourism and related topics. Also, ConTur allows an intelligent and adapted integration, to both the site where it integrates, in this case the "Portal Destino 3.0", and to the user's vision.

This is an advanced tourism Website that allows the enrichment of its contents with those produced by non-institutional travellers, hotel chains and any other agent that may have information about the destination. Given the multiplicity of possible sources of content, security mechanisms and reliability, enabled by the content manager, are of particular relevance. Among the features included are:

- Find useful information, reliable, rich and of quality, on the destinations of interest based on key concepts, instead of searching based on keywords.
- Have a centralized inquiry point of information about tourist destinations.
- Request and filter content.
- Add content and update content from other points
- Know the experience of other travellers through social networks

The portal will thus have two potential target users. On the one hand, the public administration that will use the portal for the aggregation and content management and on the other hand, the tourists, that can access the portal via the Internet to view the contents, allowing the user the preparation of their visits in a more practical way. "Portal Destino 3.0" is the technical basis for the generation of functionalities available to a fully functional Web portal, in which pages or articles that describe a tourist destination that is being offered and promoted through this infrastructure, are shown.

The objective is to be able to present the end-user, a configurable infrastructure with which he can modify the characteristics of access and display of contents offered by the manager. That is, an end user must have the ability to decide which components want, in order to obtain content, data or information, and the way the end user want them displayed. Similarly, this is a requirement for receptive managers that act as intermediaries between the obtaining of the contents of the various sources and the displaying in a particular format to the end user. These content integration capabilities are achieved through interoperable interfaces of items or services that are usually called gadgets or widgets.The experience of use of the "Portal Destino 3.0" is different from the one experienced by visitors of tourist sites today. The information that complements the destinations they are searching, come from annotated sources with semantic values. This allows an automatic system, reasoning if the results are appropriate to the user or not.

The use case "Portal Destino 3.0" in ConTur, has as its main theme, the destination Atapuerca[4] (Burgos, Spain). The use case focuses on aspects of tourism (hotels, restaurants, attractions and activities) around the archaeological site of Atapuerca, and its purpose is to provide contextualized information and of interest to the visitor, through the widgets installed in it. Widgets downloaded from Contur[5] can be installed in Portal Destino 3.0. Some of them are already developed and being tested, as beta prototypes.

5.2 Mi guia multidestino

One of the objectives of the travel agencies like Barcelo Viajes is to help customers on their travel plans, trying to supply them with much information as possible about the destinations and other offered services. Besides there is a lot of information of tourist destinations and services, sometimes, this information is old and not complete, so the information provided to the customer has a poor quality.

"Mi guia multidestino" ("My multi-destiny guide" in English) is a web application which address this requirement The main objective of the web application is to provide the functionality to organize the route of the leisure travels based on high-quality information collected from heterogeneous sources of information about destination and tourism services. From the technological basis, travel agencies request a centralized repository that offers up-to-date, integrated and high-quality information about the destinations and tourism services. Moreover, above this repository, the web application should provide functionalities and services to manage the data, users and configuration.

The solution proposed by ConTur is to offer a unique point of reliable and trusted information about the tourism resources. This tool allows search, filter and update of tourism information in an easy way and results in a cost and time reduction of the travel agencies employees. Furthermore, as the information provided to the customer is of top-quality, the relationship between agencies, stakeholders, destinations and the

---

tourist has a positive impact. Moreover, the communication interfaces of the platform are based on technological and tourism standards to handle the integration between travel agency tools and ConTur.

As indicated, the "Mi guía multidestino" web application provides the travel agencies a useful data information provider about user desires, experiences and behaviors. The results of the datawarehouse of this information will allow marketing departments of the travel agencies and destination a better effectiveness of their campaigns. "Mi guía multidestino" is part of the Barcelo Viajes Intranet provided to their business customer partners. Nowadays is at prototype and validation phase, and is predicted to be part of the main Barcelo Viajes web-site[6] in the next months.

6 Impact and Future Research

In the different market reports about the current situation of the tourism sector, one of trends identified is the growing demand of tourism contents by the main source markets. The quality, trust and variety of these contents are the base for a subsequent selection of tourism products by the tourists or final customers. Moreover, these reports also includes the shared idea that the emergence of Internet in the tourism chain-value as the technology used by the tourist to select and buy their holidays and tourism products. Internet has boosted the audience (at an international level) of the tourism products, make it easier for the tourism providers to commerce with their products. Furthermore, Internet positions itself as one of the key component in the tourism product distribution through the whole chain-value.

In the last years, several private and public entities are working in the development of new technological standards for the tourism market. This research is focused on the ecommerce through the use of some specifications based on XML messages. Some example of this research is the Travel Technology Initiative (TTI)[7] and Open Travel Alliance (OTA)[8]. In both cases, TTI and OTA, it is identified a lack of content share and data interoperability about tourism between the different actors of the market. ConTur tries to cover part of this lack and bring to the actors and easy way to share contents through its platform and open-source software and services. Because of this and taking into account some figures about the market analysis reports and its trend, it is considered that there is a niche for ConTur, to bring the contents of various service providers to consumers, where the technological solution explained in this paper is the most suitable solution to solve the market request.

7 Conclusions

In this paper the ConTur platform is presented, an intelligent content management system for the tourism sector. ConTur proposes an alternative technological solution for the current challenge of providing integrated added-value access to the content

---

[6] www.barceloviajes.com

[7] http://www.tti.org

[8] http://www.opentravel.org

information sources which allow a more trusted and easy interaction between tourist and service providers. ConTur acts as the unifying channel of contents coming from different sources (web pages, social media, public catalogues…) about tourism products and destinations. The technological solution proposed in this paper allows to filter, catalogue, aggregate, merge and integrate contents from heterogeneous sources in a intelligent, reliable and trusted way. Furthermore, two use case scenarios, "Portal Destino 3.0" and "Mi guia multidestino", demonstrates how different stakeholders, like public providers and private travel agencies, can merge and provide up-to-date information to their customers.

References

Berners-Lee, T. H. (2001). The Semantic Web. Scientific. American, May 2001.

Freeling (2010), FreeLing User Manual 2.2. Talp Research Center in Universitat Politècnica de Catalunya, Available at http://nlp.lsi.upc.edu/freeling/doc/userman/userman.pdf

Bilbao, S. (2010). E4.1 y E4.2 Modelo semántico de restricciones y requisitos de configuración. Available at http://contur.labs.andago.com

Mittal, S. and Frayman, F. (1989). Towards a Generic Model of Configuration Tasks. In Proceedings of the Ninth International Joint Conference on Artificial Intelligence, 1395–1401. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.

Murua, I. L. (2005). La web semántica aplicada al empaquetamiento dinámico de servicios y productos turísticos. Conferência IADIS Ibero-Americana WWW/Internet 2005

Piñero, R. (26 de Marzo de 2007). TechTear. Los mashups uno de los pilares de la web 2.0. Available at http://www.techtear.com/2007/03/26/los-mashups-uno-de-los-pilares-de-la-web-20

Prantner, K., Ding, Y., Luger, M., Yan, Z., Herzog, C. (2007). Tourism Ontology and Semantic Management System: Stateof-the-arts Analysis, *IADIS International ConferenceWWW/Internet 2007 (WWW/Internet2007)*, Vila Real,Portugal.

SAML (2005). Security Assertion Markup Language. OASIS Standard, Available at http://saml.xml.org/saml-specifications

W3C: World-Wide Web Consortium. Semantic Web Brief Guide. Available at http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica

Werthner, H., Grün, C., Pröll, B., Feilmayr, C., Barta, R. (2009). "Covering the Semantic Space of Tourism - An Approach based on Modularized Ontologies"; Vortrag: 6th European Semantic Web Conference (ESWC 2009), Heraklion, Griechenland; in: *ACM International Conference Proceeding Series Proceedings of the 1st Workshop on Context, Information and Ontologies*, ACM, 1, ISBN: 978-1-60558-528-4; S. 1 - 8.

X.509 (2005). The directory: Public-key and attribute certificate frameworks. ITU-T Recommendation X.509:2005 | ISO/IEC 9594-8:2005.

Acknowledgements