JAKARA: Developing Smart Dynamic Packaging Based Linked E-tourism Data

Afaf MERAZI[1]
[1]EEDIS Laboratory, Department Of Computer Science, Djillali Liabès University of Sidi Bel-Abbès, Algeria

Abstract— *E-Tourism has become the world's largest industry, composing of numerous enterprises such as airlines, hoteliers, car rentals, leisure suppliers, and travel agencies. It involves the buying and selling of tourism products and services via internet and includes many processes (e.g., dynamic packaging) which provide useful new options for consumers. Therefore, Intermediaries needs e-Business solutions to offer the growing demand of personalised products and services. Nowadays, The Semantic Web extends the existing Web, adding a multitude of language standards and software components to give humans and machines direct access to data such as GoodRelations. This paper describes a semantic e-Business platform for tourism package intermediaries that allow adaptation and customisation of tourism products and services to the desires and restrictions of the tourist at any time. We present JAKARA, a system designed to support convenient GoodRelations annotation and intelligent querying of linked Semantic Web resources oriented e-tourism Data.*

Keywords: Semantic Web, E-Tourism, GoodRelations Ontology, Tourism Products and Services, Semantic Query

1. Introduction

E-Tourism is a very successful and dynamic industry in the world business. In recent years, with many countries turning to tourism to supplement their economies, there has been a massive expansion of tourism vendor offerings. As more travel arrangements are made online, pressure is put on e-Tourism website developers to provide efficient and easy to use interfaces and intelligent services.

The Tourism industry is an information intensive business, and the amount of information is increasing rapidly; however, efficient access to this information is becoming a challenge. Due to the deluge of tourism vendor offerings and ever increasing numbers of travelers, the limitations associated with conventional e-Tourism websites are coming to the fore.

In general, e-tourism consists of travel planning and booking, which is the mainly successful business model on the World Wide Web. However, the task of planning an individual trip on the Web still consumes a time and seems to a complicated endeavor. Most of the huge number of tourism websites provides isolated and unified information about accommodation (hotel,...), transportation (flight,...), services (rental car,...), events (concert,...) or they relate that information in a very limited manner letting the consumer the heavy task of putting all the pieces together. There exists currently no integrated service for arranging personalized trips called to any desired destination, relying on distributed information sources which have to be reasonably combined.

Currently, with most tourism information systems, consumers need to visit multiple independent websites to plan their trip, register their personal information multiple times, spend hours or days waiting for response or confirmation, and make multiple payments by credit card. Consumers are discouraged by the lack of functionality. Dynamic packaging applications are emerging in response to these limitations and have caught the attention of major universal online travel agencies.

E-tourism is a perfect application area for Semantic Web technologies since information integration, dissemination, and exchange are the key backbones of the travel industry. Therefore, the Semantic Web can considerably improve e-tourism applications. The aim of this paper is to describe JAKARA system, which is designed to support convenient GoodRelations annotation and intelligent querying of Semantic Web resources. GoodRelations Annotator software is used by a website owner to generate RDF description describing the content of their website. The RDF description is essentially instance data that conforms to OWL e-tourism ontology. Query functions are facilitated by CPSPARQL query engine developed by [3]. The semantic query facility is accessed remotely via a Web-based GUI and provides the consumer with a number of search options. Once a query is submitted, a list of matching results is displayed to the consumer.

2. Dynamic Packaging

Dynamic Packaging [6]: "An industry buzzword for enabling the consumer (or booking agent) to build a customized itinerary by assembling multiple components of their choices and complete the transaction in real-time".

Dynamic Packaging is very different from prepackaged travel (see Figure 1). Prepackaged travel relies on selling to the customer a complete package that includes usually transportation, accommodations, car rental, activities and other services... These packages are made usually months

in advance and published in brochures or sold online. These packages allow the different actors of the travel industry from producers to resellers to offer "mass-market" products and to operate relatively simple business processes that allow them to have higher margins.

These "mass packages" offer: Inflexible dates, Fixed itineraries, Very limited options. But as they are made months in advance, they also often hinder the optimization of revenues through yield-management techniques that are based on adjusting price and availability to demand in real-time. In Dynamic Packaging, the process is different even if the result could seem to be the same to the end customer: here, the components are "drawn from the inventories of the travel producers and combined to satisfy a particular customer requirement, which is collected during an interactive dialog".

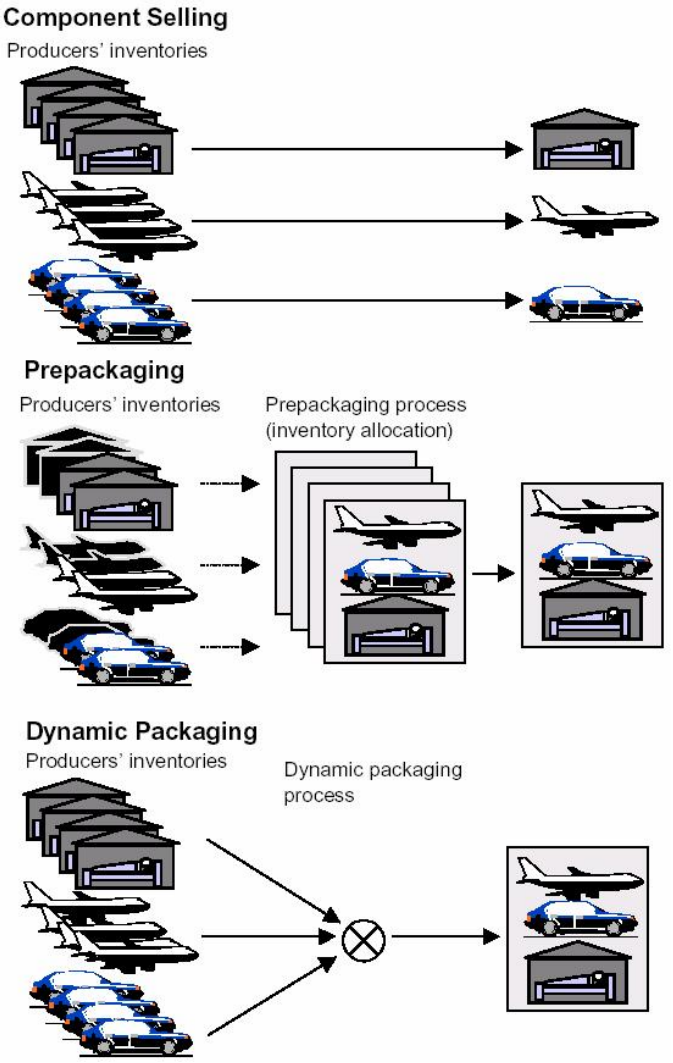

Fig. 1: Difference between dynamic packaging and component selling

Smart Dynamic Packaging is a collection of several services together in real-time according to availability, product-specific rules, especially requests (preferences) of clients (operators, tour operator). It usually includes transportation, accommodation and food, but may also include tours, excursions and other services. Any exploration of how to create smart dynamic packages must begin with an evaluation of available travel web resources; times that activities are possible and the logistics of combining different vacation components. Good questions to explore are:

- **What** can a tourist do in my area?
- **Where** are located the interesting places to see and visit?
- **When** can the tourist visit a particular place?
- **How** can the tourist get to its destination to see or do an activity?
- **What local resources** can I combine to create attractive packages?

3. Summary of E-tourism ontologies

In tourism domain, there may already exist different in-house taxonomies and catalogues which are designed and used internally by tourism agents to help them for managing heterogeneous tourism data. Efforts are made to generate global standards to facilitate inter and intra tourism information exchange (e.g., by the WTO "World Tourism Organization" which consisting of collection and collation of statistical information on international tourism). Ontologies play an important role to facilitate semantic integration of heterogeneous data. In this section, we identify several publicly available formal tourism ontologies which show the current status of the efforts and may serve as a basis for problem specific tourism ontologies (see Figure 2).

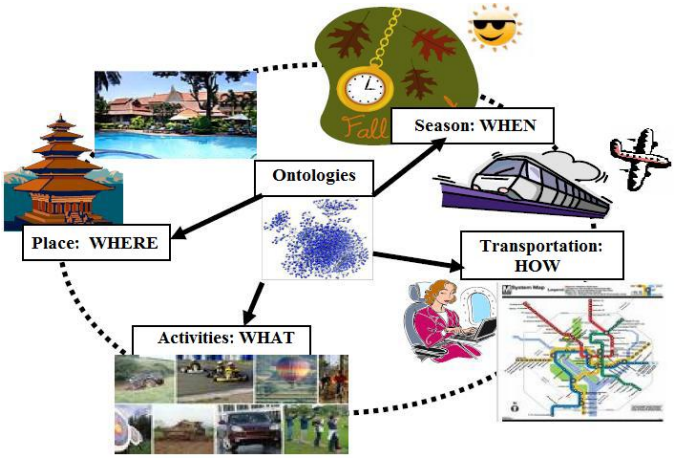

Fig. 2: The e-tourism ontologies development methodology

3.1 Harmonise Ontology

The Harmonise[1] Ontology was created within the EU Tourism Harmonisation Network (THN) established by eCTRL (E-Commerce and Tourism Research Laboratory[2]), IFITT (International Federation for IT and Travel & Tourism[3]), and others. It is specialized to address interoperability problems in the area of e-tourism focusing on data exchange. The goal of the ontology is to allow tourism organizations to exchange data and information without changing their local data structures and information systems. Harmonise is an ontology-based mediation by mapping different tourism ontologies. The used ontology language is RDF/RDFS and the ontology itself contains about 200 concepts mainly dealing with accommodation and events.

3.2 Mondeca Tourism Ontology

Mondeca[4] Tourism Ontology includes important concepts of the tourism domain which are defined in the WTO thesaurus[5] managed by the WTO (World Tourism Organization). These concepts include terms for tourism object profiling, tourism and cultural objects (place, museum, restaurant, housing, transportation, events, etc...), tourism packages and tourism multimedia content. The used ontology language is OWL and the ontology itself contains about 1000 concepts.

3.3 OnTour Ontology

The OnTour Ontology[6] is an ontology created especially for the tourism domain and was developed by DERI (Digital Enterprise Research Institute). In addition to normal tourism concepts (location, accommodation) it also includes concepts that describe leisure activities and geographic data. The used ontology language is OWL-DL[7].

3.4 OpenTravel Alliance (OTA)

The OTA (Open Travel Alliance)[8] members are organizations that represent all segments of the travel industry, along with key technology and service suppliers. The OTA Specification defines XML Message Sets packages that contain about 140 XML Schema documents corresponding to events and activities in various travel sectors (booking, availability checking, reservation and cancelling of flights, hotels and rental cars).

[1]http://www.harmonise.org
[2]http://ertrl.itc.it:8080/home/index.jsp
[3]http://www.ifitt.org
[4]www.mondeca.com
[5]The WTO Thesaurus includes information and definitions of the topic tourism and leisure activities. www.world-tourism.org
[6]http://ontour.deri.org/ontology/ontour-02.owl
[7]http://e-tourism.deri.at/ont/index.html
[8]http://www.opentravel.org/

3.5 Other Ontologies

A Tourism Ontology[9] developed by the University of Karlsruhe contains four different sub-ontologies for the tourism domain defining about 300 concepts and more than 100 relations. The EON Travelling Ontology[10] is mainly designed for the travel domain developed by INA (Institut National de l'Audiovisuel - France). The TAGA Travel Ontology[11] is another travel focused ontology that provides typical concepts of travelling combined with concepts describing typical tourism activities such as itineraries, customers, travel services, and service reservations.

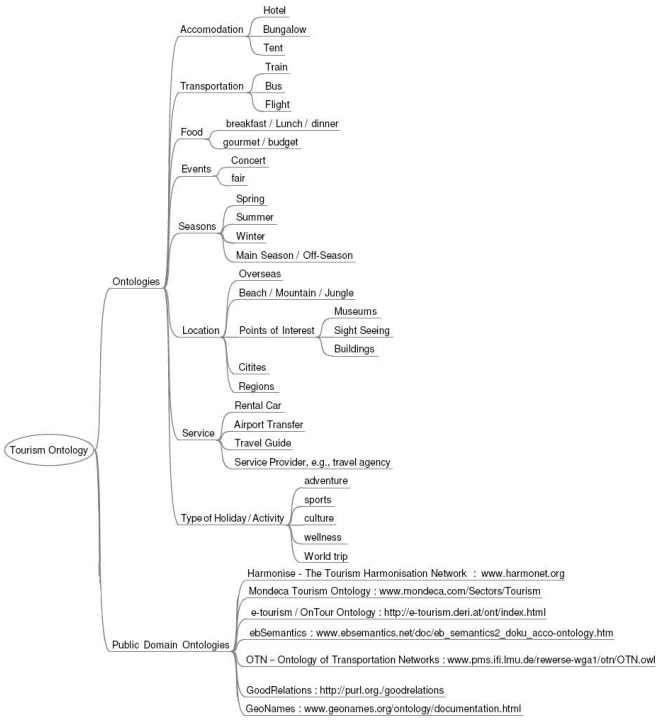

Fig. 3: Tourism ontology dimensions

The following code is showing an example of e-tourism ontology which includes classes Accommodations, Guestroom, etc. and where the Accommodations classes have object property called hasRoom.

```
<owl:Ontology rdf:about=""/>
<owl:Class rdf:ID="Accomodation">
<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty>
<owl:ObjectProperty rdf:ID="hasRoom"/>
</owl:onProperty>
<owl:someValuesFrom>
<owl:Class rdf:ID="Guestroom"/>
</owl:someValuesFrom>
</owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf....
```

[9]http://ontobroker.semanticweb.org/ontos/comparing.html
[10]http://opales.ina.fr/public/ontologies/EON-TravellingOntology-v0.2.daml
[11]http://taga.sourceforge.net/owl/travel.owl

4. GoodRelations: An Ontology for Describing Web Offers

Semantic Web + E-Commerce = GoodRelations

Representing the semantics of data is a need that has recently emerged from the Semantic Web area which intends to bring meaning to the unprecedented and tremendous amount of data published over the Web. Tim Berners-Lee presents the Semantic Web [5] as "not a separate Web, but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation". He also conditions its success to an automated reasoning: "computers must have access to a structured collections of information and sets of inference rules that they can use to conduct automated reasoning [7]". The Semantic Web is based on Ontologies - languages to formally describe domain specific concepts and their relations - coupled to reasoning engines, to perform domain fact deductions. A promising application domain for Semantic Web technology is the annotation of products and services offers on the Web so that consumers and enterprises can search for suitable suppliers using products and services ontologies. Martin Hepp introduces GoodRelations Ontology to allow for each company who possesses a website to describe and represent all their products and services offering for intelligent search and deterministic response End user. We give an overview of some definitions about this domain application:

"The GoodRelations ontology[12] provides a conceptual model for a consolidated view on commerce data on the Web, e.g.: companies, store locations, offers, product descriptions, pricing, payment, shipment, and warranty information. It is provides the vocabulary for annotating e-commerce offerings (see Figure 4)." [9]

"The GoodRelations ontology provides a generic yet lightweight vocabulary for describing in a machine-readable way the details of offers made on the Web. This allows vendors to encode their offers so that Semantic Web search engines can find such Web resources precisely. It empowers them to return exactly matching offers for a given need."

"GoodRelations is a lightweight yet sophisticated vocabulary that allows manufacturers and show operators to express the exact meaning of their offers made on the Web in a machine-readable way. This empowers search engines to support more precise search, and partners in the value chain to automate their content integration tasks. (1) to sell, lease, repair, dispose, and maintain commodity products and (2) to provide commodity services. GoodRelations allow describing the relationship between (1) Web resources, (2) offerings made by those Web resources, (3) legal entities, (4) prices, (5) terms and conditions, and the aforementioned ontologies for products and services[13] (6)." [8]

In the following, we define an overview of the relevant conceptual entities by giving an informal definition and additional explanations in the Table 1.

Table 1: The most important conceptual elements of GoodRelations.

Feature	Description	Reference
Business Entity.	present the company Web site (mailing address and contact details).	*gr:BusinessEntity*
Products and Offerings.	describe the actual products or services that are being offered, and the offering, the actual business function (sell, repair, dispose, etc.) and other commercial properties.	*gr:ProductOrServiceModel*
Eligible Customers and Regions.	The types of customers for which the given Offering is valid.	*gr:eligibleCustomerTypes*
Price Specifications.	information on the prices fixed by vendors.	*gr:hasPriceSpecification*
Delivery Options and Delivery Charge Specifications.	shipment charge.	*gr:availableDeliveryMethods*
Payment Options and Payment Charge Specifications.	procedure for transferring the monetary amount for a purchase	*gr:acceptedPaymentMethods*
Warranty Promises.	offer includes a bundle of services in case of defects or malfunction.	*gr:hasWarrantyPromise*
Bundles.	placeholder instance for unknown instances of a mass-produces commodity.	*gr:ProductOrServicesSome InstancesPlaceholder*
Services and Value Ranges.	numerical interval that represents the range of a certain quantitative Product or Service Property in terms of the lower and upper bounds for one particular Product Or Service.	*gr:QuantitativeValueFloat gr:TypeAndQuantityNode*
Shop Locations and Opening Hours.	availability of a particular Offering at or from and defines the opening hours for a given DayOfWeek.	*gr:availableAtOrFrom gr:hasOpeningHoursSpecification*

Every object called resource in RDF has a global identifier in the form of a URI. The design of Linked Data [4] provides recommendations on how to publish both kinds of resources, and it is also an initiative on publishing data from various closed databases in RDF. Linked e-tourism Data is

[12] http://www.heppnetz.de/projects/goodrelations/primer/20080808/

[13] We are talking about real services, not Semantic Web services or Web services.

an example of such a Linked Data database of information coming from a set of GoodRelations ontologies including people, places, organizations, agencies, etc. Each of them is one non information resource with a unique URI and its information representation. It is recommended to reuse resources from such knowledge bases instead of creating new individuals.

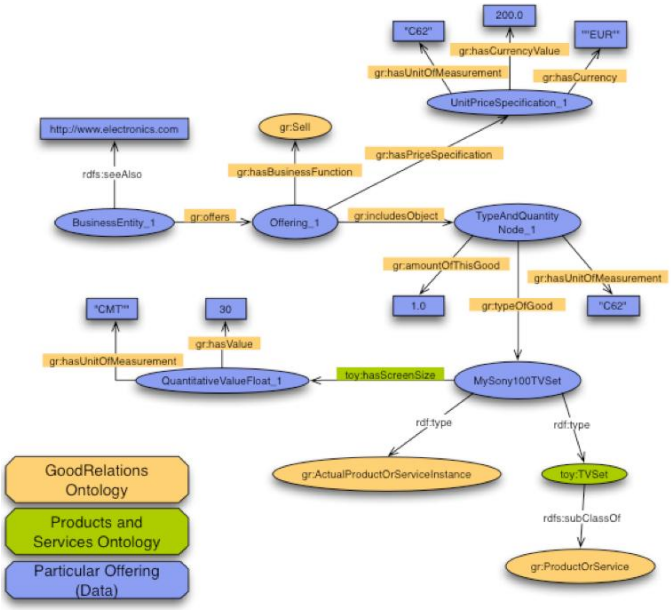

Fig. 4: RDF graph of the minimal example

5. Framework Architecture & Overview

Nowadays, E-tourism is a leading industry in e-business. Furthermore, it is viewed as information intensive industry with a long value chain where information plays an important role for decision and action making. There are many challenges to solve in the e-tourism distribution; the most important is to enable interoperability between information systems, allowing seamless information exchange between tourism organizations. Semantic technologies provide methods and concepts facilitating integration of tourism information originating from various sources on top of so-called ontologies (formal domain conceptualisations). Semantic technologies offer a significant potential for better cross-system integration and a more versatile linkage of available multimedia tourism data based on ontology references and other types of semantic mark-up (such as geo-referencing of data).

Generally, The interoperability problem between cooperating systems can be differentiated into:

1) **Information-level interoperability**, which addresses clashes between different data representations and their meaning, and
2) **Service-level interoperability**, when different processes should cooperate to automatically perform business between enterprises.

Our framework concentrates at the information level interoperability problem and proposes solution for this level only. Information heterogeneity between different systems can be categorized as follows:

- Semantic clashes: addressing different meaning of concepts of different systems, and
- Representational clashes: caused by different data representations formats With the respect to the impact on the data exchange, structuring conflicts can be differentiated into:
- Fully mappable: all clashes can be resolved without any loss of information, and
- Partially or nonmappable: covering the structural conflicts for which any conceivable transformation will cause a loss of information.

In this paper, we analyze the complexity of product description in the tourism distribution and propose to use the GoodRelations ontology that covers the representational needs of typical e-commerce scenarios in the commodity segment. In this section, we develop the requirements on a e-tourism products and commodity services offered on the Web.

- Due to the heterogeneity of the e-tourism product, *tourism agency consultant or a person who is planning the trip itself must have access to multiple sources of information.*
- Most e-tourism products (e.g., hotel rooms, flight tickets) are time-constrained and can't be stocked/stored. To support this, *basic products must have well-defined interfaces with respect to consumer needs, prices or distribution channels and an abstract model of the product (e.g., its description).*
- E-tourism product is also immaterial, meaning that traveler cannot see or touch the tourism product before the trip. *That is why reliable information about destination, accommodation options and other parts of the tourism product is extremely important for both people working in tourism industry and tourists themselves. Tourism product cannot be stored in storage*.
- If a hotel room or a seat in an airplane remains empty today, this is lost revenue for the tourism company. This is a reason why *effective distribution and inventory management are key factors* in the tourism business.

The purpose of JAKARA is to create a system that provides a tangible benefit over existing e-tourism Web portals by allowing tourism customers to search the underlying products or services of a website, thus producing results that more closely match the customer's needs and preferences denoting "smart dynamic packaging". This is

achieved by using GoodRelations ontology "a promising application domain for Semantic Web technology" to infer knowledge about resorts and seamlessly integrating that knowledge so that it can be used by a tourism customer when searching for suitable travel.

The system is limited to the GoodRelations'annotation and querying of a specific Country travel & tourism web-sites. Only websites annotated consistent with the e-tourism ontology (OnTour Ontology viewed in section 3.3) employed are included in the system. Such techniques are recognized as important for the integration of tourism & travel information, but fall outside the scope of what the system aims to demonstrate. Web-page annotations are conceptually consistent products and services offers instance data of the e-tourism ontology, and are queried by the GUI using a database looks-up from a Jena back-end knowledge base. We present the percentage of websites that contain at least a sufficient amount of information in the various categories is visualized in Figure 5.

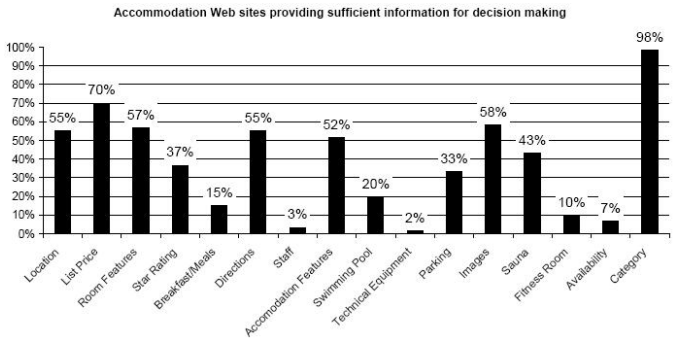

Fig. 5: Percentage of e-tourism Web sites

The JAKARA architecture (see Figure 6) is designed to support convenient annotation based product and service offers application domain and intelligent querying of Semantic Web resources.

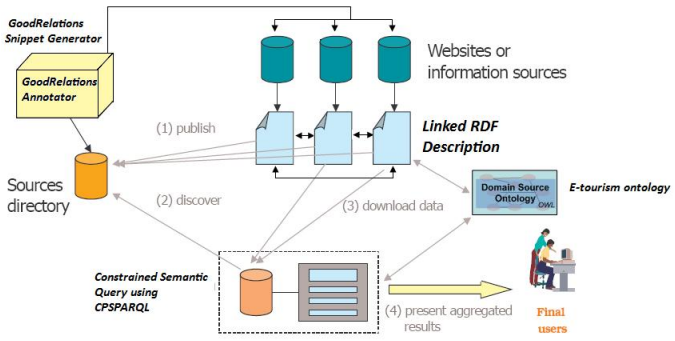

Fig. 6: JAKARA Architecture

5.1 GoodRelations Annotator based e-tourism ontology

Annotation software is used by a website owner to generate RDF descriptions describing the content of their website, it is a form-based tool that helps owners of any business in the world produces a basic yet detailed description of its range of products or services called GoodRelations Annotator [10] "Inspired by the impact of the FOAF-a-Matic tool on the availability of FOAF data on the Web, we developed a form-based tool that can be used by any business in the world to create a basic yet detailed description of its range of products, payment and delivery options, store locations, opening hours, and eligible customer types and eligible regions[14] Internally, it uses a HTML form with a substantial amount of JavaScript for validating the input, and a Python server-side component that handles the conversion and returns RDF/XML (in the future also RDFa snippets [1]) plus instructions on how to publish the data."

The RDF description is essentially instance data that conforms to an OWL e-tourism ontology. Query functions are facilitated by a Jena based CPSPARQL[15] query engine that uses a Pellet reasoner and the linked GoodRelations ontologies to infer knowledge about the e-tourism domain. The query facility is accessed remotely via a Web-based GUI and provides the consumer. Once a query is submitted, a list of matching results is displayed to the consumer.

5.2 Querying GoodRelations offers using CPSPARQL

"SPARQL [11] can be used to express queries across diverse data sources, whether the data are stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be result sets or RDF graphs." There are several proposed extension to SPARQL:

- PSPARQL [2] which allows query graph patterns involving regular expressions. This provides a lot of flexibility in expressing graph patterns that can be matched. However, PSPARQL supports only a pattern matching query paradigm.
- CPSPARQL [3] that further extends (P)SPARQL by allowing, for example, complex constraints on nodes and edges of traversed paths.

[14]The tool is available at: http://www.ebusiness-unibw.org/tools/goodrelations-annotator/

[15]available at: http://psparql.inrialpes.fr/

We decide to implement the CPSPARQL query language to describe the preferences and needs of the consumer furthermore to create a Smart dynamic Packaging which must correspond to products and services offering requirement. In the following, we give an example in CPSPARQL on how to query respective product and services data on the Semantic Web. For example, the price of each direct trip is no more than 500:

```
PREFIX gr: <http://purl.org/goodrelations/v1#>
PREFIX ex: <http://www.heppnetz.de/ontologies/goodrelations
/examples#>
SELECT ?offering ?uri ?maxprice ?currency
WHERE { CONSTRAINT const1 ]ALL ?Trip]:
?offering gr:includesObject ?TypeAndQuantityNode .
?TypeAndQuantityNode gr:typeOfGood ?something .
?something rdf:type ex:trip .
?offering gr:hasBusinessFunction gr:Sell.
?offering rdfs:seeAlso ?uri .
?offering gr:hasPriceSpecification ?priceSpecification .
?priceSpecification rdf:type gr:UnitPriceSpecification .
?priceSpecification gr:hasCurrency ?currency .
?priceSpecification gr:hasMaxCurrencyValue ?maxprice .
FILTER (?maxprice < 500)
}
```

6. Conclusion

The Semantic Web tools and technologies are useful in developing the next generation of e-Tourism websites. Some of the benefits would be, more flexible searching of appropriate websites, less precise input from users in requesting what they want and a standard way of packaging intelligence for a suite of e-Tourism applications. Some limitations in applying the Semantic Web are due to the nature of the Internet as it stands today. For example, how much automation is it safe to apply on the Internet, given that security remain a problem?. This paper presented an overview of the Semantic Web, and described in detail a framework for developing e-Tourism applications based on the technologies of the Semantic Web.

Hence, one big challenge to successfully develop dynamic packaging applications is finding a solution to cope and integrate the nonstandard way of defining e-tourism products and services. We therefore suggested Web services and semantics as emerging technologies that can be used to deal with the lack of standard and enable data intention. These two technologies have already been recognized in the Semantic Web and can considerably improve e-tourism systems.

References

[1] B. Adida, M. Birbeck, S. McCarron, and S. Pemberton, *RDFa in XHTML: Syntax and Processing Ű A collection of attributes and processing rules for extending XHTML to support RDF*, W3C Recommendation, http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014/, 2008.

[2] F. Alkhateeb, J. F. Baget, and J. Euzenat, *Extending SPARQL with regular expression patterns (for querying RDF), Journal of Web Semantics*, 7(2):57–73, Elsevier Science Publishers B. V. 2009.

[3] F. Alkhateeb, J. F. Baget, and J. Euzenat, *Constrained regular expressions in SPARQL*, in: Hamid R. Arabnia and Andy Marsh (eds), Proc. international conference on semantic web and web services (SWWS), Las Vegas (NV US), pp 91–99, 2008.

[4] T. Berners-Lee, *Linked Data*, in Design Issues, http://www.w3.org/DesignIssues/LinkedData.html, 2006.

[5] T. Berners-Lee, J. Hendler, and L. Ora, *The Semantic Web*, Scientific American, 2001.

[6] J. Cardoso and C. Lange, *A Framework for Assessing Strategies and Technologies for Dynamic Packaging Applications*. Journal of Information Technology & Tourism, 9(1), 27–44, 2007.

[7] A. Gaignard, and J. Montagnat, *Survey on semantic data stores and reasoning engines* VIP Project - ANR-09-COSI-03 - Milestone 1.2.1 pp 1–31. Version 1.1, September 14, 2010.

[8] M. Hepp. *GoodRelations: An Ontology for Describing Products and Services Offers on the Web*, Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management (EKAW2008), September 29 - October 3, 2008 (forthcoming), Acitrezza, Italy, Springer LNCS, Vol. 5268, pp. 329–346, 2008.

[9] M. Hepp. *Product Variety, Consumer Preferences, and Web Technology: Can the Web of Data Reduce Price Competition and Increase Customer Satisfaction?*, in: T. Di Noia and F. Buccafurri (Eds.): EC-Web 2009, LNCS 5692, p. 144. Springer-Verlag, Berlin Heidelberg, 2009.

[10] M. Hepp, A. Radinger, A. Wechselberger, A. Stolz, D. Bingel, T. Irmscher, M. Mattern, and T. Ostheim, *GoodRelations Tools and Applications*, in: Poster and Demo Proceedings of the 8th International Semantic Web Conference (ISWC 2009), Washington, DC, USA, October 25–29, 2009.

[11] E. Prud'hommeaux and A. Seaborne (ed.), *SPARQL Query Language for RDF*. http://www.w3.org/TR/rdf-sparql-query/, 2008.