e-Review of Tourism Research (eRTR), Vol. 10, No. 2, 2012
http://ertr.tamu.edu

Special Issue – ENTER 2012 Idea Exchange

Improving Tourism Information Search and Retrieval

Sonia Bilbao, Jesús Herrero

Tecnalia Research & Innovation, Spain
[sonia.bilbao]@tecnalia.com

Abstract

The research described in this paper tries to bridge the existing gap between the tourism domain and user queries support in travel planning. The search engine proposed is based on a semantic model which allows defining a query language that is flexible and robust to changes in the tourism data model. This is critical as new tourism products, services and experiences are constantly emerging and the information needed to describe them varies considerably from one item to another. The search engine supports specifying customizable constraints and two aspects generally not considered in current travel planning online sites: nested sorting of the results and possibility to customize the amount and type of information to be retrieved about a tourism resource.

Keywords: Search engine; Information retrieval; Ontology; Travel guide configuration.

1 Introduction

The amount and variety of tourism products and services offered online continues to grow. Tourism is moving from a commodities industry (flights, hotels, car rentals) to an experience industry (enjoy the sunset while trying the local cuisine, drive a Ferrari at some of the best racetracks…). Besides, it is extremely difficult to predict tourist's demand. Even the same traveller is interested in different products and services depending on the motivation and features of the trip, elements of the external environment and the particularities and attributes of destinations [Buhalis, 2010]. Hence, with such a huge amount of information available online and traveller segmentation not being a trivial task, it becomes increasingly important to offer configurable search engines that help travellers find products that match their demand and expectations. However, current search engines for travel planning offer limited query support and there is a substantial gap between the language representations in tourism related Web pages and user queries [Xiang, 2009] [Xiang, 2008].

2 Research objectives

The main goal of the research was to design a search engine that would improve information search and retrieval, i.e. finding the information you need quickly and accurately. To this end, the research looked into the following topics:

- Design a robust query language against changes in the data model structure.
- The query language should require minimum technical skills
- Search engine should allow customizable filters for content
- Nested sorting of the results
- Customize the amount and type of information to be retrieved about a resource

3 Methodology and Technology Description

The following figure shows how the retrieval process works. Firstly, the requirements and constraints [Hentenryck, 2009] of the end user are formalized into a search query that describes the entities that can be part of the solution we are looking for. Secondly, the sorting criteria describe how these entities will be ranked in the list of results provided by the search engine. Thirdly, the retrieval values describe the kind of information the end user wants to obtain about each entity, e.g. name and address about an accommodation; price, description and location about an experience; timetable about a transportation, and so on.

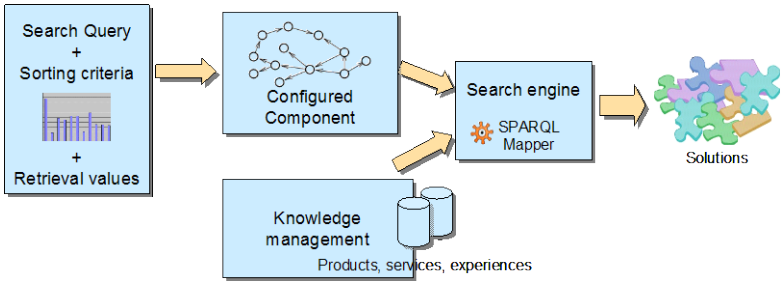

Figure 1 Steps of the retrieval process

This information, i.e. search query, sorting criteria and retrieval values, is written in terms of the semantic model described in chapter 3.1 and as an instance of the concept ConfiguredComponent. The search engine receives this instance as input and uses the SPARQL Mapper to generate the SPARQL query which, once executed by the knowledge management API, will return the results we are looking for.

3.1 High level view of the semantic model

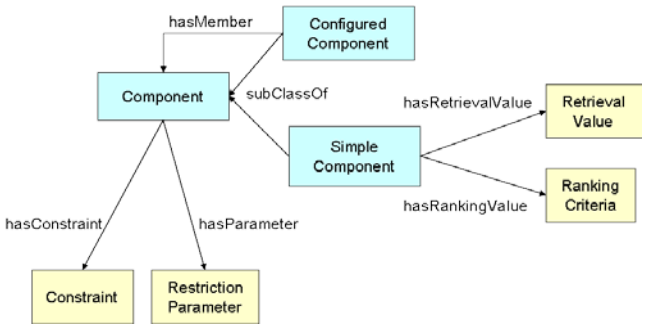

Figure 2 Configuration model. Main concepts and properties

The semantic model (see figure 2) used as input to the search engine is an evolution and adaptation of the configuration ontology defined in the OBELIX project [1] [Altuna et al, 2004] to the tourism sector's requirements. Ontologies were used for modelling purposes in order to benefit from semantic reasoning capabilities and to seamlessly allow for sharing and reuse of knowledge in computational form.

3.2 Loose coupling of queries and data model

New tourism experiences are constantly emerging and the information needed to describe them varies considerably from one case to another. For this reason, the query language used to retrieve the information from the search engine should be robust against changes in the data model structure. Otherwise, the solution would not be flexible enough to adapt to future market demands and trends.

In the semantic model, the entities to be searched can be of two types: simple or configured components. A simple component is the most basic entity that participates in a configuration design whereas a configured component is an assembly of items (simple and/or configured). Examples of simple components from the tourism domain are accommodations, flights, car-rental services and restaurants whereas configured components could be holiday packages, experiences or a travel guide. Each simple component is loosely associated to a domain concept by the property "describedByConcept" that links the item to its URI (Uniform Resource Identifier) or local name. This way, the identifier of the item is not hardcoded in the engine's software and can be easily modified in a text file without the need of technical skills.

[1] Ontology-based electronic integration of complex products and value chains.http://cordis.europa.eu/fetch?CALLER=PROJ_ICT&ACTION=D&CAT=PROJ&RCN=61165 (FP5 project)

3.3 Customizable filters for content

The filters that can be applied to the content to be retrieved are classified under two categories: restriction parameters and constraints. By restriction parameters we understand filters to the group of results that match the requirements. They are used to specify global restrictions. Two have been defined: "NumberOfComponents" (to restrict the maximum number of items of a given type that can be part of the results) and "Offset" (to specify the interval of results to be returned). By constraints we understand filters to intrinsic characteristics that describe an item and that can differentiate two items of the same class, e.g. number of stars of a hotel. Constraints enumerate the possible values that the properties of an item may take.

3.4 Nested sorting of the results

When the search engine returns a large number of results, it becomes necessary to have mechanisms for ranking these results based on relevancy factors. The relevance of these parameters varies depending on the traveller and the context. Currently available travel search engines usually have the limitation that they only allow to define at most one ranking criterion. However, the semantic model defined covers the definition of nested ranking criteria based on three sorting algorithms:

- Natural Order Sorting: used to sort the results in an ascending or descending ordering, i.e. if the results are strings they are sorted in alphabetical order and in the case of quantities a natural order numerical sorting algorithm is used.
- Group Sorting: used to define subsets of the results in terms of property values. E.g. sort accommodations by type displaying first bed and breakfast results, next apartments and finally hotels.
- Interval Sorting: used to define subsets of the results in terms of ranges of numerical property values.

3.5 Customize the amount and type of information to be retrieved about a resource

When searching for a product or service we are not always interested in the same type of information. This depends not only on the user but also on his/her context. Sometimes we want to ask for the popularity of an accommodation. Other times we are interested in images displaying the hotel's amenities. Just before the trip it is how to get there what is really of concern for us. However, current tourism search engines do not allow customizing the results and do not provide tailored information. The system developed allows configuring at run-time the information to be returned. This information is grouped under two categories: simple retrieval values, that represent key-value pairs of information and complex retrieval values, which are used to return nested elements that represent the same piece of information.

4 Business Case Description

The main objective of the business case was to provide users with an online tool that would allow them creating their own travel guides and sharing these guides with their friends on the social network. In order to use the tool the user has to log in. This way, the system can make suggestions based on the user's profile and preferences and on the items previously consulted. The next step in the guide creation process is to select a name for the guide, the travelling dates and the type of access for the guide, i.e. private, if only the owner will have access to the guide; public, if anyone will have access to the guide; and protected, if the user's friends on the social network will have access to the guide. Next, the user can add destinations to include in the guide and for each destination the user can decide the type of tourism items to be discovered by the search engine. The search query includes explicit information provided by the traveller but can also include implicit information reasoned from the system's knowledge configuration [Viappiani, 2006] or from the user's context [Cheverst et al, 2000]. The output of the search engine will be products, services and experiences that could be of interest to the user. For each item, the user can read related information such as name and description, location in a map, price, popularity, comments and multimedia content (photos or videos). The item can be included in the travel guide and all the information can be consulted at any time.

5 Conclusions

The search engine described in the paper tries to improve the information search and retrieval processes involved in travel planning. It was successfully tested in a travel guide creation process of a travel agency site.

As the input is an instance of an ontology, it can be easily generated either automatically from an online form or manually using an ontology editor such as Protégé requiring minimum technical skills.

The query language is designed to be flexible and robust to changes in the data model. If new tourism items are considered or existing ones are modified, the input to the search engine, i.e. selection and ranking criteria or retrieval values, can be easily modified at run-time without affecting the system's software. Hence, the system will work as expected without modifications even if the model containing the products and services is modified.

The semantic model supports customizable filters that can be restriction parameters if they specify global restrictions, or constraints if they apply to intrinsic characteristics that describe an item.

Finally, the search engine supports two aspects generally not considered in current travel planning online sites: nested sorting of the results and possibility to customize the amount and type of information to be retrieved about a resource.

References

Altuna, A., Cabrerizo, A., Laresgoiti, I., Peña, N. and Sastre, D. (2004). Co-operative and distributed configuration. In Net.ObjectDays (NODe), pages 69–80, Erfurt. Germany.

Buhalis, D., Michopoulou, E., G. (2010). Information-Enabled Tourism Destination Marketing: Addressing the Accessibility Market, Current Issues in Tourism, vol 10.6.

Cheverst, K., Davies, N., Mitchell, K., Friday, A. and Efstratiou, C. (2000). Developing a context-aware electronic tourist guide: some issues and experiences. In CHI 2000: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 17-24.

Hentenryck, P. V. and Michel, L. (2009). Constraint-Based Local Search. The MIT Press.

Viappiani, P., Faltings, B. and Pu, P. (2006) 'Evaluating preference-based search tools: a tale of two approaches', in 22th National Conference on Artificial Intelligence(AAAI), Boston, USA, pp. 205-211.

Xiang, Z., U. Gretzel, and D. R Fesenmaier. (2009). Semantic representatin of the online tourism domain. Journal of Travel Research, 47(4):440-453.

Xiang, Z. (2008). Assessing the Online Tourism Domain from an Information Search Perspective. Ph.D. Dissertation. Temple University, Philadelphia, PA, USA. Advisor(s) Daniel R. Fesenmaier. AAI3300384.